



In the Name of God

The 10th Seminar on Probability and
Stochastic Processes

The Proceeding of Refereed and Invited Papers

August 19-20, 2015

Yazd University

Yazd, Iran

Preface

This volume contains the full style of contributed articles at the 10th Seminar on Probability and Stochastic Processes (SPSP10) in Iran. The Seminar on Probability and Stochastic Processes is a two days event held every two years with cooperation of the Iranian Statistical Society. The SPSP10 organized jointly by Yazd University and the Iranian Statistics Society, is taking place from 19th to 20th of August at Yazd University. The organizing committee of the SPSP10 warmly welcomes the participants to the historical city of Yazd, hoping that their stay in Yazd will be happy and fruitful one. More than 300 participants have taken part in this seminar. Out of 310 articles, 106 articles are selected for oral presentation and 122 articles are selected for poster presentation by the scientific committee of the seminar. We have made every effort to make the seminar as worthwhile as possible.

Acknowledgement

We wish to express our thanks to all whose help has made this gathering possible. In particular we would like to express our gratitude to the administration of Yazd University and Iranian Statistical Society. We gratefully acknowledge the careful efforts of our staff in the Department of Statistics at Yazd University, specially Dr R. Roozegar, Dr A. A. Jafari and Miss R. Meshkati. We also wish to thank all organizations for their great educational as well as financial support in holding SPSP10.

Organizer Committee
August 2015

Organizing Committee

Owlia M. S. (University President)

Pezeshk, H.

Zakerzadeh, H. (Chairman)

Almodaresi, S. M. T.

Gholam Nezhad, J.

Barid Loghmani, Gh.

Morovati Sharifabadi, A.

Dolati, A. (Scientific Chair)

Mirhosseini, S. M. (Executive Chair)

Shahzadeh Fazeli, A.

Scientific Committee

Dolati A.	Yazd University
Amini M.	Ferdowsi University of Mashhad
Torabi H.	Yazd University
Chinipardaz R.	Shahid Chamran University of Ahvaz
Soltani A. R.	Shiraz and Kuwait Universities
Zohouri Zanganeh B.	Sharif University of Technology
Alamatsaz M. H.	Isfahan University
Farnoosh R.	Iran University of Science and Technology
Mahmoudi E.	Yazd University

Table of Contents

Stochastic comparisons of Harris family	1
<i>Somayeh Abbasi, Mohammad Hossein Alamatsaz</i>	
Stochastic Orderings of the Extended Generalized Exponential Distribution	10
<i>Maliheh Abbasnejad</i>	
Estimation of stress-strength reliability based on generalized Rayleigh records	19
<i>Mousa Abdi, Akbar Asgharzadeh</i>	
Wavelet Shrinkage Estimator Of Regression Function Based on Maximum A Posterior	28
<i>Mahmoud Afshari, Behzade Gholizadeh</i>	
On the Entropy Rate of Stationary Markov Chains	37
<i>Mansour Aghababaei Jazi, Hasan Zarei</i>	
On Distribution of the Sum of Dependent Random Variables	46
<i>Najmeh Ahmadi, Ali Dolati</i>	
A Bayesian method for bandwidth choice with length-biased data	55
<i>Masoud Ajami bakhtyarvand</i>	
Survey to Gene Network Interaction in Microarray Data Analysis	64
<i>Hamid Alavi majd, Atefeh Talebi, Ali Dolati, Kambiz Gilani</i>	
An asymmetric Markov switching GARCH model: estimation and forecasting	71
<i>Seyede Nafiseh Alemohammad, Saied Rezakhah, Sasan Hosseinalizadeh</i>	
A Sufficient Condition for Absolute Continuity of the Distribution of Lévy Processes	80
<i>Kasra Alishahi, Erfan Salavati</i>	
An alternative to general record models: dynamic after-selection estimation	86
<i>Morteza Amini, Nader Nematollahi</i>	

Truncated unified skew-elliptical distributions with application to order statistics	95
<i>Mehdi Amiri, Ahad Jamalizadeh, Seyyed Roohollah Roozegar</i>	
Distribution of a random sum of exponential random variables	104
<i>A. Asgharzadeh, M. Habibi</i>	
Ordering Results for Aggregate Claim Amounts from Two Heterogeneous Marshall-Olkin Extended Weibull Portfolios and their Applications in Insurance Analysis	113
<i>Ghobad Barmalzan, Abedin Haidari, Amir T. Payandeh Najafabadi</i>	
On the analysis of the two parameter exponential distribution based on progressive type II censored data	122
<i>Reza Arabi Belaghi, Hossein Bevrani, Mojtaba Mohammadi</i>	
A method for selecting generating OWA operator weights models	131
<i>Alireza Chaji</i>	
A Note on the Parrondo's Paradox	140
<i>N. Ejlali, H. Pezeshk, and M. Sadeghi</i>	
A normal-Poisson distribution: Theory and applications	148
<i>Fatemeh Esfandiyari, Eisa Mahmoudi</i>	
American Contingent Claims with Extra Information for the Buyer	157
<i>Neda Esmaeeli, Peter Imkeller</i>	
Reliability analysis of a semi-Markov repairable system	163
<i>Malek Fathizadeh, Omid Kharazmi</i>	
Improving in hybrid Monte Carlo via Halton sequence with increasing and decreasing informity .	170
<i>Behrouz Fathi-Vajargah, Hamzeh Torabi, Hojatollah Zakerzadeh, Ali Asghar Latifi Zadeh</i>	
On lower bounds for the variance of functions of random variables with weighted distributions ..	179
<i>Faranak Goodarzi, Mohammad Amini, Golam Reza Mohtashami Borzadaran</i>	

Usual Stochastic Order Between Largest Claim Amounts and its Applications in Insurance	189
<i>Abedin Haidari, Ghobad Barmalzan and Amir T. Payandeh Najafabadi</i>	
Periodically correlated space-time autoregressive Hilbertian processes of order one	199
<i>Maryam Hashemi, Jorge Mateu</i>	
The inactivity time of the $(n - k + 1)$ -out-of- n system with exchangeable components under stress-strength models	206
<i>Shahrokh Hashemi, Ebrahim Salehi</i>	
Adaptive Progressively Type-II Censored Competing Risks Data	215
<i>Fariba Hemmati, Esmaeil Khorram</i>	
Fully Stochastic Single-Period Mean-Variance Portfolio Optimization	225
<i>Reza Keykhaei, Bardia Panahbehagh</i>	
Asymptotic Coverage Levels of Confidence Intervals for prediction in Functional Linear Regression	233
<i>Omid Khademnoe, S. Mohammad E. Hosseini-Nasab</i>	
On the modified of generalized weighted exponential distribution	242
<i>Omid Kharazmi</i>	
A New Bayesian Unit Root Test in Unobserved-ARCH Models	250
<i>Fazlolah Lak, Mahmood Afshari, Behzad Gholizadeh</i>	
α -power transformation method for generating distributions	259
<i>Abbas Mahdavi</i>	
A goodness-of-fit test for the extreme value distribution in the presence of outliers	267
<i>Abbas Mahdavi, Hossein Negarestani</i>	
A note on Bayesian inference for the Topp-Leone distribution	276
<i>Mahdi Mahdizadeh</i>	

A gamma-Weibull distribution: Model, properties and applications	283
<i>Rahmat Sadat Meshkat, Hamzeh Torabi</i>	
The Cowan's Copula and its Extension	292
<i>Seyyed Mohsen Mirhosseini, Ali Dolati</i>	
Longitude Shape Analysis by Using the Spherical Coordinates	298
<i>Meisam Moghimbeygi, Mousa Golalizadeh</i>	
Prediction in GINAR(1) Time Series Model	307
<i>Mehrnaz Mohammadpour, Khadije Rezvani</i>	
Notes on Generalized Reversed Failure Rate	317
<i>G. R. Mohtashami Borzadaran</i>	
Randomized order restricted estimators for treatment means with a control group	327
<i>Reza Momeni, Javad Etminan</i>	
Asymptotic behavior of least square estimator for stochastic differential equation arisen from RL electrical circuits	335
<i>P. Nabati</i>	
Complete Convergence for Weighted sums of FGM random sequences	344
<i>H. Naderi, M. Amini and A. Bozorgnia</i>	
A High Order Finite Difference Scheme for Stochastic Advection Diffusion Equations	358
<i>Mehran Namjoo, A. Mohebbian</i>	
The discrete beta generalized exponential distribution	367
<i>Vahid Nekoukhrou</i>	
Some Results on APND Random Variables	376
<i>Hamid Reza Nili Sani, Mohammad Amini, Abolghasem Bozorgnia</i>	

Some Asymptotic Results on Improved LASSO Estimators	384
<i>Mina Norouzirad, Mohammad Arashi, A.K.Md. Ehsanes Saleh</i>	
Stochastic Differential Equations for SIS and SIR epidemic models	393
<i>Mahmood Parsamanesh, Rahman Farnoosh</i>	
An Approximation to Finite- and Infinite-time Ruin Probabilities of Compound Poisson Processes	401
<i>Amir T. Payandeh Najafabadi</i>	
Slashed generalized normal distribution: Model and properties	415
<i>Atefe Pourchitsaz, Eisa Mahmoudi, Hamide Jafari</i>	
A note on the mean remaining time to failure of the censored data under type-I and type-II censoring schemes	423
<i>Mohammad Hossein Poursaeed</i>	
A Long Memory Linear Arch Model and R/S Statistic for Estimating its Long Memory Parameter	433
<i>Rohollah Ramezani, Seyed Abolfazl Mousavi</i>	
Periodically Correlated Multi-component Locally Stationary Processes	440
<i>S. Rezakhah, N. Modarresi</i>	
Prediction via bivariate concomitants of bivariate order statistics	454
<i>Seyyed Roohollah Roozegar, Ahad Jamalizadeh, Mehdi Amiri</i>	
A study for testing Zero inflated Negative Binomial versus Negative Binomial distribution	462
<i>Zahra Saberi, Maryam Hassanpour Khademi</i>	
The residual lifetime of linear consecutive k -out-of- n systems with exchangeable components	469
<i>Ebrahim Salehi</i>	
Expressions for the mean of the order statistics from the skew-normal distribution and their application	477
<i>Mahdi Salehi, Mahdi Doostparast</i>	

Dirichlet processes applied to clustering of longitudinal data via piecewise mixed-effects models	485
<i>Elaheh Shams, Reyhaneh Rikhtehgaran</i>	
Hill-Type Estimators and Invariance	493
<i>Mehdi Shams, Gholamreza Hesamian</i>	
A New Bivariate Exponential distribution	501
<i>Zahra Sharifonassabi, Mohammad Hossein Alamatsaz, Iraj Kazemi</i>	
Expected Number of Real Roots of Certain Gaussian Random Trigonometric Polynomials	509
<i>Soudabeh Shemehsavar</i>	
Discrete Exponential INAR(1) model	518
<i>Masume Shirozhan, Mehrnaz Mohammadpour</i>	
A new measure of inaccuracy based on past lifetime	526
<i>Raziyeh Siroosi, Mohammad Khorashadizadeh</i>	
On bivariate dynamic cumulative past entropy	534
<i>Raziyeh Siroosi, Mohammad Khorashadizadeh</i>	
Generalized Mixture Representations for Inactivity Times of Coherent Systems	542
<i>Abdurrahim Sotoudeh, Sareh Goliforushani</i>	
Stable Multiple Markov Processes	549
<i>Mehrdad Taghipour, Ali reza Nematollahi</i>	
Weighted Integration to Solutions of SDE's with uniformly elliptic diffusion	557
<i>Mahdieh Tahmasebi</i>	
Information Measures for Weighted Distributions of Exp-G Family	562
<i>Saeid Tahmasebi, Ali Akbar Jafari</i>	
Extreme Value Distributions for Skew-t Distributions	570

Asma Taimouri, Mohsen Khosravi, Ahad Jamalizadeh

Hypothesis Testing of Hidden Periodicities by Quantile Periodogram 578

Marziye Tazikinejad, Mahnaz Khalafi, Majid Azimmohseni

Goodness-of-Fit Test for Normality Based on Empirical Distribution Function 585

Hamzeh Torabi, Narges Montazeri Hedesh

Prediction For Lindley Distribution Based on Progressively Type-II Censored Samples 593

R. Valiollahi, A. Asgharzadeh

A new generalized Pareto - Poission distribution generated by beta random variables 601

Shahram Yaghoobzadeh Shahrastani

Comparing Recurrent and Vector Singular Spectrum Analysis with Artificial Neural Networks in Time series Analysis 609

Masuod Yarmohammadi

Application of Statistical Tests of Anisotropy for Fractional Brownian Textures to Mammograms 616

Gholamhossein Yari, Soodeh Moallemian

Exponential Pareto II Distribution 624

Gholamhossein Yari, Zahra Tondpour

Shrinkage estimation of scale parameters for two exponential distributions based on record values 632

Hojjatolah Zakerzadeh, Mahdieh Karimi

Rate of Convergence for Sample Covariance Operator of H -valued Periodically Correlated Time Series .. 640

Atefeh Zamani

Skew Normal Autoregressive Models with Markov Property 648

Hamid Reza Zareifard, Fardin Malek Mohammadi

Test of independence for some well-known bivariate distributions 655

Mansoor Zargar, Hadi Jabbari and Mohammad Amini

The Topp-Leone Distribution: Revisited 664

Ahmad Pourdardvish, S.M.T.K. MirMostafae, Kimia Naderi



Stochastic comparisons of Harris family

S. Abbasi ¹, M. H. Alamatsaz ²

^{1,2}Department of Statistics, University of Isfahan, Isfahan

Abstract: Harris family of distributions is a known class of extended distributions of the lifetime of a series system with variable number of components. In this paper, we show that several stochastic orderings are preserved by transformation to Harris family. Our results happen to enfold several previous findings regarding the Marshall-Olkin family in this connection.

Keywords Ageing intensity ordering, Harris family, Marshall-Olkin distribution, Proportional stochastic ordering

1 Introduction

There are two main reasons for studying semiparametric families. One is that they provide extended models with higher flexibility in data applications. Second is that they provide better understanding of old family of distributions. In this regard, by adding new parameters to certain family of distributions, Marshall and Olkin (1997) and Aly and Benkherouf (2011) introduced two large classes of flexible distributions, called Marshall-Olkin and Harris families of distributions, respectively. The method that Marshall-Olkin and Harris generated their family of distributions is as follows: Let Y_1, Y_2, \dots be a sequence of independent and identical (*iid*) random variables (r.v.) with common survival function $\bar{F}(\cdot)$. Suppose $X = \min\{Y_1, Y_2, \dots, Y_N\}$, where N is a r.v. independent of Y_i that take positive integer values with probability generating function (pgf)

$$P_N(t) = E(t^N) = \sum_{n=0}^{\infty} t^n P(N = n).$$

That is, X can be considered as the lifetime of a series system with iid component lifetimes Y_1, Y_2, \dots, Y_N when the number of components, N , is itself a r.v. independent of Y_i .

¹Speaker's Name : Somayeh Abbasi s_abbasi_29@yahoo.com

It is seen that the survival function $\bar{H}(\cdot)$ of X can be written as $\bar{H}(x) = \sum_{n=0}^{\infty} [\bar{F}(x)]^n P(N = n)$ and thus

$$\bar{H}(x) = P_N(\bar{F}(x)). \quad (1.1)$$

By replacing $P_N(\cdot)$ with the geometric pgf in Eq.(1.1), Marshall-Olkin distribution was introduced by Marshall and Olkin (1997). They, then, in the same paper generated the Marshall-Olkin extended exponential (MOEE) and Marshall-Olkin extended Weibull (MOEW) r.v.'s, respectively. Aly and Benkherouf (2011) replaced Harris pgf

$$P_N(s; \theta, k) = \left\{ \frac{\theta s^k}{1 - \theta s^k} \right\}^{1/k}, \quad k > 0, \quad 0 < \theta < 1, \quad \bar{\theta} = 1 - \theta, \quad (1.2)$$

introduced by Harris (1948), into Eq.(1.1) and defined the Harris family of survival functions \bar{H} as

$$\bar{H}(x; \theta, k) = \left(\frac{\theta \bar{F}^k(x)}{1 - \theta \bar{F}^k(x)} \right)^{1/k}, \quad k > 0, \quad 0 < \theta < \infty \quad \bar{\theta} = 1 - \theta. \quad (1.3)$$

Note that here $\theta > 0$, i.e., it is not restricted the interval (0,1). $F(x)$ in (1.3) is called the baseline distribution function (df.) and θ is called the tilt parameter. It is easily seen that hazard rates corresponding to $F(x)$ and $H(x; \theta, k)$, namely, $r_F(\cdot) = \frac{f(\cdot)}{F(\cdot)}$ and $r_H(\cdot; \theta, k) = \frac{h(\cdot; \theta, k)}{H(\cdot; \theta, k)}$, are related by

$$r_H(x; \theta, k) = \frac{r_F(x)}{1 - \theta \bar{F}^k(x)}; \quad -\infty < x < \infty, \quad 0 < \theta < \infty \quad k > 0. \quad (1.4)$$

Clearly, $r_H(x; \theta, k)$ is shifted below ($\theta \geq 1$) or above ($0 < \theta \leq 1$) $r_F(x)$. When $k = 1$, pgf (1.2) reduces to the positive geometric pgf and we arrive at the Marshall-Olkin distribution.

Recently, several results were provided by Batsidis and lemonte (2014) in connection with the behavior of the failure rate function for Harris family.

Our aim is to investigate preservation of stochastic orderings by transformation to Harris family. Stochastic orderings are important tools for comparing probability distributions. In this paper, in Section 2, we shall review various types of stochastic orderings used in the sequel. In Section 3, it is observed that certain stochastic orderings of the baseline family are preserved by transformation to Harris family with the same tilt parameter.

2 Stochastic Ordering

We use various types of stochastic orderings such as usual, shifted and proportional stochastic orderings. In below we show acronyms which are used in this paper. For more details, we refer to Lillo et al. (2001), Belzunce et al. (2002), Shaked and Shantikumar (2007), Marshall and Olkin (2007) and Jarrahiferiz et al. (2013).

ACRONYMS AND ABBREVIATIONS

st	simple stochastic	mrl	mean residual life
lr	likelihood ratio	plr	proportional likelihood ratio
hr	hazard rate	phr	proportional hazard rate
rh	reversed hazard rate	prh	proportional reversed hazard rate
cx	convex	c	convex transform
LOR	log odds ratio	disp	dispersive
su	super additive	AI	ageing intensity
lr \uparrow	up likelihood ratio	lr \downarrow	down likelihood ratio
hr \uparrow	up hazard rate	hr \downarrow	down hazard rate
rhr \uparrow	up reversed hazard rate	rhr \downarrow	down reversed hazard rate

In Table 1, we summarize some useful relationships among the above mentioned stochastic orderings to be used in the sequel.

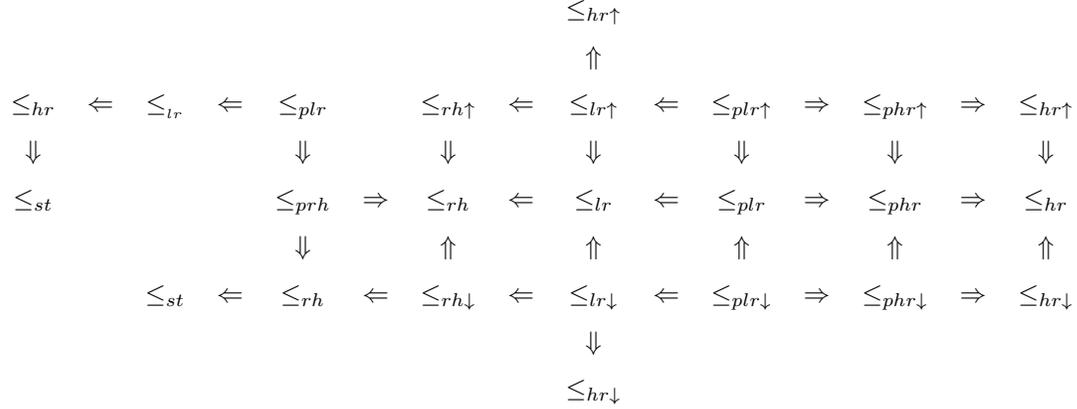
3 Preservation properties

Assume that the baseline df., $F(x)$, in (1.3) is absolutely continuous with pdf $f(x)$. Then, pdf and df. associated with $\bar{H}(x; \theta, k)$ in (1.3) are given by

$$h(x; \theta, k) = \frac{\theta^{\frac{1}{k}} f(x)}{(1 - \theta \bar{F}^k(x))^{1 + \frac{1}{k}}}; \quad -\infty < x < \infty, \quad 0 < \theta < \infty, \quad k > 0, \quad (3.1)$$

$$H(x; \theta, k) = 1 - \left[\frac{\theta \bar{F}^k(x)}{(1 - \theta \bar{F}^k(x))} \right]^{\frac{1}{k}}; \quad -\infty < x < \infty, \quad 0 < \theta < \infty, \quad k > 0, \quad (3.2)$$

Table 1: Some useful relations among various types of stochastic orderings



respectively.

Let X_1 and X_2 be two r.v.'s with df.'s F_1 and F_2 and pdf's f_1 and f_2 , respectively. Suppose that Y_1 and Y_2 are their corresponding Harris r.v.'s, i.e., those with baseline df.'s F_1 and F_2 , respectively. In this section, we shall study stochastic orderings preservation of the baseline distribution by its corresponding Harris distribution.

Marshall and olkin (2007) have shown that the simple stochastic, hazard rate, convex transform and star orderings are preserved by transformation to Marshall-Olkin family. In what follows, their results are generalized for any $k > 0$ in Eq.(1.3).

Theorem 3.1. *Simple stochastic ordering is preserved by transformation to Harris family.*

Proof. Let $X_1 \leq_{st} X_2$, i.e., $\bar{F}_1(x) \leq \bar{F}_2(x)$ for all x . Thus, for all $k > 0$ we have $\bar{F}_1^k(x) \leq \bar{F}_2^k(x)$ and consequently, for any θ , $\bar{F}_1^k(x) - \theta \bar{F}_1^k(x) \bar{F}_2^k(x) \leq \bar{F}_2^k(x) - \theta \bar{F}_1^k(x) \bar{F}_2^k(x)$. This is equivalent to $\bar{F}_1^k(x)[1 - \theta \bar{F}_2^k(x)] \leq \bar{F}_2^k(x)[1 - \theta \bar{F}_1^k(x)]$ which implies $[\frac{\theta \bar{F}_1^k(x)}{1 - \theta \bar{F}_1^k(x)}]^{\frac{1}{k}} \leq [\frac{\theta \bar{F}_2^k(x)}{1 - \theta \bar{F}_2^k(x)}]^{\frac{1}{k}}$. This proves the assertion. \square

Theorem 3.2. *If $\theta > 1$, then*

- i) up hazard rate ordering is preserved by transformation to Harris family.*
- ii) proportional hazard rate ordering is preserved by transformation to Harris family.*
- iii) hazard rate ordering is preserved by transformation to Harris family.*

Proof. We proof the (i) part. Proofs of other parts are similar and thus omitted.

Let $X_1 \leq_{hr\uparrow} X_2$. So, for all x and $t \geq 0$ we have $r_{F_1}(x+t) \geq r_{F_2}(x)$. Since hazard rate ordering is implied by up hazard rate ordering and simple stochastic ordering is implied by hazard rate ordering (Table 1), for any x and all $k > 0$ we have $\bar{F}_1^k(x) \leq \bar{F}_2^k(x)$. But survival function is decreasing, so for all $t \geq 0$, we have $\bar{F}_1^k(x+t) \leq \bar{F}_1^k(x) \leq \bar{F}_2^k(x)$. Thus, when $\theta > 1$, we have $-\theta \bar{F}_1^k(x+t) \leq -\theta \bar{F}_2^k(x) \implies 1 - \theta \bar{F}_1^k(x+t) \leq 1 - \theta \bar{F}_2^k(x)$. Thus, by Eq.(1.4), the result follows. \square

The following example shows that up hazard rate ordering is not preserved by transformation to Harris family, when $0 < \theta < 1$.

Example 3.3. Let X_1 and X_2 be two r.v.s having Erlang distribution with survival functions $\bar{F}_1(x) = (1+2x)e^{-2x}$, $\bar{F}_2(x) = (x+1)e^{-x}$ and hazard rates $r_{F_1}(x) = \frac{4x}{1+2x}$, $r_{F_2}(x) = \frac{x}{x+1}$, for $x > 0$, respectively. So, $X_1 \leq_{hr\uparrow} X_2$. However, it is seen that for some $0 < \theta < 1$, $t > 0$ and some $x > 0$, $r_{H_1}(x+t; \theta, k) \not\geq r_{H_2}(x; \theta, k)$ or, equivalently, $\frac{\bar{H}_2(x; \theta, k)}{\bar{H}_1(x+t; \theta, k)}$ is not increasing in x , i.e., up hazard rate ordering is not preserved by transformation to Harris family, when $0 < \theta < 1$.

Corollary 3.4. Let X_1 and X_2 be two r.v.s with mrl functions m and l and Harris r.v.s Y_1 and Y_2 , respectively. Suppose that $\frac{m(t)}{l(t)}$ increases in t . If $X_1 \leq_{mrl} X_2$, then $Y_1 \leq_{mrl} Y_2$.

Proof. By Theorem 2.A.2 of Shaked and Shanthikumar (2007), the assertion follows because if $X_1 \leq_{mrl} X_2$ and $\frac{m(t)}{l(t)}$ increases in t , then $X_1 \leq_{hr} X_2$. Thus, by Theorem 3.2(i) we can conclude that $Y_1 \leq_{hr} Y_2$. But by sufficiency of hazard rate ordering for mrl ordering (Shaked and Shanthikumar (2007)), this implies that $Y_1 \leq_{mrl} Y_2$. \square

For the spacial case when $k = 1$, preservation of reversed hazard rate ordering by transformation to Marshall-Olkin family is investigated in the following theorem.

Theorem 3.5. If $0 < \theta < 1$, then

- i) up reversed hazard rate ordering is preserved by transformation to Marshall-Olkin family.
- ii) reversed hazard rate ordering is preserved by transformation to Marshall-Olkin family.

Proof. i) Let $X_1 \leq_{rh\uparrow} X_2$. So, for all x and $t \geq 0$ we have, $\tilde{r}_{F_1}(x+t) < \tilde{r}_{F_2}(x)$. Since reversed hazard rate ordering is implied by up reversed hazard rate ordering (Table 1) and simple stochastic

ordering is implied by reversed hazard rate ordering (Table 1), we have $\bar{F}_1(x) \leq \bar{F}_2(x)$. But survival function is decreasing, thus, $\bar{F}_1(x+t) \leq \bar{F}_1(x) \leq \bar{F}_2(x)$. If $0 < \theta < 1$, then $\bar{\theta}\bar{F}_1(x+t) \leq \bar{\theta}\bar{F}_2(x)$, or $1 - \bar{\theta}\bar{F}_1(x+t) \geq 1 - \bar{\theta}\bar{F}_2(x)$ which yields $\frac{1}{1-\bar{\theta}\bar{F}_1(x+t)} \leq \frac{1}{1-\bar{\theta}\bar{F}_2(x)}$. This proves the assertion. For proof of (ii) it is sufficient to put $t=0$ in the above proof. \square

The following example shows that reversed hazard rate ordering and up reversed hazard rate ordering are not preserved by transformation to Marshall-Olkin family, when $\theta > 1$.

Example 3.6. Let X_1 and X_2 be two exponential r.v.'s with hazard rates 8 and 4, respectively. So, $X_1 \leq_{rh} X_2$ and $X_1 \leq_{rh\uparrow} X_2$. However, for some $\theta > 1$ and some $x > 0$, $\tilde{r}_{H_1}(x; \theta, 1) \not\leq \tilde{r}_{H_2}(x; \theta, 1)$ or, equivalently, $\frac{H_2(x; \theta, 1)}{H_1(x; \theta, 1)}$ is not increasing in x for some $\theta > 1$, i.e., reversed hazard rate ordering is not preserved by transformation to Marshall-Olkin family when $\theta > 1$. Moreover, it is seen that for some $\theta > 1$ and some $x > 0$, $\tilde{r}_{H_1}(x+t; \theta, 1) \not\leq \tilde{r}_{H_2}(x; \theta, 1)$ or, equivalently, $\frac{H_2(x; \theta, 1)}{H_1(x+t; \theta, 1)}$ is not increasing in x for some $\theta > 1$, i.e., up reversed hazard rate ordering is not preserved by transformation to Marshall-Olkin family when $\theta > 1$.

Remark 3.7. Note that for the special case when $k = 1$, the log-odds function of a r.v. X is equal to the log-odds function of the corresponding Harris r.v. Y . Consequently, the log-odds ratio ordering is also preserved by transformation to Marshall-Olkin family.

For the aging intensity ordering, we have the following.

Theorem 3.8. Let X_1 and X_2 be non-negative r.v.'s. For all $k > 0$, if $X_1 \leq_{AI} X_2$ and $X_1 \leq_{hr} X_2$, then $Y_1 \leq_{AI} Y_2$ provided that $\theta > 1$.

Proof. Let $k > 0$ and $\theta > 1$. $Y_1 \leq_{AI} Y_2$ if, and only if,

$$\frac{1}{r_{H_1}(x; \theta, k)} \int_0^x r_{H_1}(u; \theta, k) du \leq \frac{1}{r_{H_2}(x; \theta, k)} \int_0^x r_{H_2}(u; \theta, k) du$$

or,

$$\frac{1 - \bar{\theta}\bar{F}_1^k(x)}{r_{F_1}(x)} \int_0^x r_{H_1}(u; \theta, k) du \leq \frac{1 - \bar{\theta}\bar{F}_2^k(x)}{r_{F_2}(x)} \int_0^x r_{H_2}(u; \theta, k) du.$$

But, we have

$$\int_0^x r_H(u; \theta, k) du = -\ln \bar{H}(x; \theta, k)$$

$$= -\ln \bar{F}(x) + \frac{1}{k} \ln\left(\frac{1 - \bar{\theta} \bar{F}^k(x)}{\theta}\right).$$

So, we should show that

$$(1 - \bar{\theta} \bar{F}_1^k(x)) \left[\frac{-\ln \bar{F}_1(x)}{r_{F_1}(x)} + \frac{1}{k} \frac{\ln\left(\frac{1 - \bar{\theta} \bar{F}_1^k(x)}{\theta}\right)}{r_{F_1}(x)} \right] \leq (1 - \bar{\theta} \bar{F}_2^k(x)) \left[\frac{-\ln \bar{F}_2(x)}{r_{F_2}(x)} + \frac{1}{k} \frac{\ln\left(\frac{1 - \bar{\theta} \bar{F}_2^k(x)}{\theta}\right)}{r_{F_2}(x)} \right]. \quad (3.3)$$

We also have $X_1 \leq_{AI} X_2$ if, and only if,

$$\begin{aligned} \frac{1}{r_{F_1}(x)} \int_0^x r_{F_1}(u) du &\leq \frac{1}{r_{F_2}(x)} \int_0^x r_{F_2}(u) du \\ \frac{1}{r_{F_1}(x)} \int \frac{f_1(x)}{\bar{F}_1(u)} du &\leq \frac{1}{r_{F_2}(x)} \int \frac{f_2(u)}{\bar{F}_2(u)} du. \end{aligned}$$

Equivalently, we have

$$\frac{-\ln \bar{F}_1(x)}{r_{F_1}(x)} \leq \frac{-\ln \bar{F}_2(x)}{r_{F_2}(x)}. \quad (3.4)$$

On the other hand, if $X_1 \leq_{hr} X_2$, for all x we have $\frac{1}{r_{F_1}(x)} \leq \frac{1}{r_{F_2}(x)}$ and also $X_1 \leq_{st} X_2$ thus, $\bar{F}_1^k(x) \leq \bar{F}_2^k(x)$. So, since $\theta > 1$ we have $\frac{1 - \bar{\theta} \bar{F}_1^k(x)}{\theta} \leq \frac{1 - \bar{\theta} \bar{F}_2^k(x)}{\theta}$. Hence, we can conclude that

$$\frac{\ln\left(\frac{1 - \bar{\theta} \bar{F}_1^k(x)}{\theta}\right)}{r_{F_1}(x)} \leq \frac{\ln\left(\frac{1 - \bar{\theta} \bar{F}_2^k(x)}{\theta}\right)}{r_{F_2}(x)}. \quad (3.5)$$

Now, adding up inequalities (3.4) and (3.5) and multiplying the left side by $(1 - \bar{\theta} \bar{F}^k(x))$ and the right side by $(1 - \bar{\theta} \bar{F}_2^k(x))$, inequality (3.3) is obtained. This completes the proof. \square

In the next lemma we need inverses of df. and survival function of a Harris distribution. It is easy to verify that Eq.(1.3) and Eq.(3.2) lead to:

$$\bar{H}^{-1}(p; \theta, k) = \bar{F}^{-1}\left(\frac{p^k}{\theta + \theta p^k}\right)^{\frac{1}{k}} \quad 0 < p < 1, \quad (3.6)$$

and

$$H^{-1}(p; \theta, k) = F^{-1}\left(1 - \left[\frac{(1-p)^k}{\theta + \bar{\theta}(1-p)^k}\right]^{\frac{1}{k}}\right) \quad 0 < p < 1. \quad (3.7)$$

Lemma 3.9. *If*

$$\bar{H}_i(x; \theta, k) = \left(\frac{\theta \bar{F}_i^k(x)}{1 - \bar{\theta} \bar{F}_i^k(x)}\right)^{\frac{1}{k}}; \quad i = 1, 2, \quad (3.8)$$

then, $H_2^{-1}(H_1(x)) = F_2^{-1}(F_1(x))$; for all x .

Proof. This result can be obtained using the assumed form of H_1 together with H_2^{-1} , which is obtainable from Eq.(3.7). Because for any $k > 0$ and $\theta > 0$

$$H_2^{-1}(H_1(x)) = F_2^{-1}\left(1 - \left[\frac{(1 - H_1(x))^k}{\theta + \bar{\theta}(1 - H_1(x))^k}\right]^{\frac{1}{k}}\right), \quad (3.9)$$

substituting $H_1(x)$ of Eq.(3.2) into Eq.(3.9) the Lemma follows. \square

Without any restriction on the tilt parameter θ , we have:

Theorem 3.10. *The following orderings are preserved by transformation to Harris family.*

i) Convex transform ordering, ii) star ordering, iii) supper additive ordering, iv) dispersive ordering.

Proof. We proof the first part. Proofs of the other parts is similar and are omitted. i) $X_1 \leq_c X_2$ is equivalent to $F_2^{-1}F_1(x)$ is convex in $x \in S_{X_1}$. Thus, by Lemma 3.9, $H_2^{-1}(H_1(x))$ is convex in $x \in S_{Y_1}$. So $Y_1 \leq_c Y_2$. \square

Corollary 3.11. *If $X_1 \leq_{Lorenz} X_2$, then $Y_1 \leq_{Lorenz} Y_2$ provided that $\frac{F_2^{-1}(x)}{F_1^{-1}(x)}$ is increasing.*

Proof. If $\frac{F_2^{-1}(x)}{F_1^{-1}(x)}$ is increasing for all $x > 0$ then, $\frac{F_2^{-1}F_1(x)}{x}$ is increasing for all $x > 0$. Thus, $X_1 \leq_* X_2$ and since Lorenz ordering is implied by star ordering (Shaked and Shanthikumar (2007)) i.e., $X_1 \leq_{Lorenz} X_2$. Since, by Theorem 3.10, star ordering is preserved by transformation to Harris family, we have $Y_1 \leq_* Y_2$. Which yields Lorenz ordering. As required. \square

References

- Aly, EAA. and Benkherouf L. (2011). A new family of distributions based on probability generating functions. *Sankhya B*, **73**, 70-80.
- Batsidis, A and Lemonte, A.J. (2014). On the Harris extended family of distributions. *Statistics*, DOI:10.1080/02331888.2014.969732
- Belzunce, F., Ruiz, J. M. and Ruiz, M. C. (2002). On preservation of some shifted and proportional orders by systems. *Statistics and Probability Letters*, **60**, 141-154.

- Harris, TE. (1948). Branching processes. *Ann Math Stat*, **19**, 474-494.
- Jarahiferiz, J., Mohtashami Borzadaran, G. R. and Rezaei Roknabadi, A. H. (2013). Some properties and applications of shifted proportional stochastic orders. *Journal of the Turkish Statistical Association*, **6**, 80-91.
- Lillo, R. E., Nanda, A. K. and Shaked, M. (2001). Preservation of some likelihood ratio stochastic orders by order statistics. *Statistics and Probability Letters*, **51**, 111-119.
- Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, **84(3)**, 641-652.
- Marshall, A. W. and Olkin, I. (2007). *Life distributions: Structure of nonparametric, semiparametric, and parametric families*. New York: Springer, Springer Series in Statistics.
- Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Academic Press, New York.



Stochastic orderings of the extended generalized exponential distribution

Maliheh Abbasnejad¹

Department of Statistics, Hakim Sabzevari University

Abstract: Recently a new distribution, named as extended generalized exponential distribution has been introduced by Kundu and Gupta (2011). In this paper investigate some famous orderings of the extended generalized exponential distribution. Also the expressions of various uncertainty measures of EGE distribution are presented and entropy orderings are discussed.

Keywords likelihood ratio ordering, dispersion ordering, hazard rate ordering, entropy.

Mathematics Subject Classification (2010): 94A17 62E99 62M99.

1 Introduction

The two-parameter generalized exponential (GE) distribution has a probability density function (pdf) and a distribution function (cdf) as follows:

$$\begin{aligned}f(x; \alpha, \lambda) &= \alpha\lambda(1 - e^{-\lambda x})^{\alpha-1}e^{-\lambda x}, \quad x > 0, \\F(x; \alpha, \lambda) &= (1 - e^{-\lambda x})^\alpha,\end{aligned}$$

where $\alpha > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. The GE distribution has been proposed by Gupta and Kundu (1999) as an alternative to gamma and Weibull distributions. The hazard rate of this distribution is an increasing or a decreasing function depending on the shape parameter. The GE distribution has been studied extensively by Gupta and Kundu (2001a, 2001b, 2007), Raqab (2002), Raqab and Madi (2005), Baklizi (2008), Kundu and Gupta (2008), Madi and Raqab (2009) and Chen and Lio (2010). One of the major disadvantages of the GE distribution is

¹Maliheh Abbasnejad : m.abbasnejad@hsu.ac.ir

that it can not be used to analyze a data set with non-monotone hazard functions, similar to Weibull or gamma distributions.

Recently Kundu and Gupta (2011) introduced a new family of distribution functions as an extension of the GE distribution with an additional shape parameter. The four-parameter extended generalized exponential (EGE) distribution has the distribution function as

$$F(x; \alpha, \beta, \mu, \lambda) = \begin{cases} \left(1 - (1 - \beta\lambda(x - \mu))^{\frac{1}{\beta}}\right)^\alpha & \text{if } \beta \neq 0 \\ (1 - e^{-\lambda(x-\mu)})^\alpha & \text{if } \beta = 0, \end{cases} \quad (1.1)$$

where $\alpha > 0$ and $-\infty < \beta < \infty$ are the shape parameters, $-\infty < \mu < \infty$ is the location parameter and $\lambda > 0$ is the scale parameter. The pdf of EGE distribution becomes

$$f(x; \alpha, \beta, \mu, \lambda) = \begin{cases} \alpha\lambda \left(1 - (1 - \beta\lambda(x - \mu))^{\frac{1}{\beta}}\right)^{\alpha-1} (1 - \beta\lambda(x - \mu))^{\frac{1}{\beta}-1} & \text{if } \beta \neq 0 \\ \alpha\lambda (1 - e^{-\lambda(x-\mu)})^{\alpha-1} e^{-\lambda(x-\mu)} & \text{if } \beta = 0, \end{cases} \quad (1.2)$$

where $\mu < x < \infty$ for $\beta \leq 0$ and $\mu < x < \mu + 1/(\beta\lambda)$ for $\beta > 0$.

Many well known distributions can be obtained as special cases of the EGE distribution. If $\beta = 0$, EGE reduces to GE; $\beta = 0$, $\alpha = 1$, EGE reduces to exponential; $\beta = 1$, $\alpha = 1$, EGE reduces to uniform; $\alpha = 1$, EGE reduces to generalized Pareto and $\beta < 0$, $\alpha = 1$, EGE reduces to Pareto.

The hazard function of EGE distribution is unimodal if $\beta < 0$, $\alpha > 1$, a decreasing function if $\beta < 0$, $\alpha < 1$, an increasing function if $\beta > 0$, $\alpha > 1$ and a bathtub shaped if $\beta > 0$, $\alpha < 1$.

In the literature several concepts of partial ordering among random variables have been considered. These concepts are useful in modeling for reliability and economics applications. For example, these concepts have been used to design better systems. For more applications see Barlow and Proschan (1975) and Fagioli and Pellerey (1993). New orderings have been proposed based on measures of entropy. For example, Ebrahimi and Pellerey (1995) used the residual shannon entropy to define a partial ordering of random variables. For orderings based on other measures of entropy, one can refer to Ebrahimi *et al.* (1999), Nanda and Paul (2006), Abbasnejad and Arghami (2008), Abbasnejad *et al.* (2010), Abbasnejad (2011) and Zardashat (2015).

In this paper we investigate some famous orderings as well as orderings based on entropy measures in EGE family. The rest of the paper is organized as follows. In section 2, we give important orderings of EGE distribution. In Section 3, we present the expressions of entropy measures for EGE family and then give entropy orderings of these distributions.

2 Some famous orderings

In this section, we provide some results on well-known orderings of EGE distributions. We need the following definition in which X and Y denote random variables with distribution functions F and G , density functions f and g and survival functions \bar{F} and \bar{G} .

Definition 2.1. *The random variable X is said to be less than or equal to Y in the*

- *Dispersion ordering, denoted by $X \stackrel{dis}{\leq} Y$, if and only if*

$$F^{-1}(v) - F^{-1}(u) \leq G^{-1}(v) - G^{-1}(u) \quad \forall \quad 0 < u \leq v < 1,$$

or equivalently $G^{-1}(u) - F^{-1}(u)$ is an increasing function of $u \in (0, 1)$.

- *Likelihood ratio ordering, denoted by $X \stackrel{lr}{\leq} Y$, if $\frac{g(x)}{f(x)}$ is increasing in x .*
- *Stochastic ordering, denoted by $X \stackrel{st}{\leq} Y$, if $\bar{F}(x) \leq \bar{G}(x)$ for all x .*
- *Hazard rate ordering, denoted by $X \stackrel{hr}{\leq} Y$, if $\lambda_F(t) \geq \lambda_G(t)$ for all $t \geq 0$.*
- *Mean residual lifetime ordering, denoted by $X \stackrel{mrl}{\leq} Y$, if $m_X(t) \leq m_Y(t)$ for all $t > 0$, where $m_X(t)$ and $m_Y(t)$ are mean residual lifetime functions of X and Y respectively and $m_X(t)$ is given by $m_X(t) = E(X - t | X > t) = \frac{\int_t^{+\infty} \bar{F}(x) dx}{F(t)}$.*
- *Convex transform ordering, denoted by $X \stackrel{c}{\leq} Y$, if $G^{-1}(F(x))$ is a convex function.*
- *Star shaped ordering, denoted by $X \stackrel{*}{\leq} Y$, if $\frac{G^{-1}(F(x))}{x}$ is increasing in x .*

Remark 2.2. *The following implications are well-known $X \stackrel{lr}{\leq} Y \Rightarrow X \stackrel{hr}{\leq} Y \Rightarrow X \stackrel{st}{\leq} Y$ and $X \stackrel{hr}{\leq} Y \Rightarrow X \stackrel{mrl}{\leq} Y$. Also $X \stackrel{dis}{\leq} Y \Rightarrow X \stackrel{st}{\leq} Y$ and $X \stackrel{c}{\leq} Y \Rightarrow X \stackrel{*}{\leq} Y$. For more details one can refer to Shaked and Shantikumar (1994).*

In the following Theorems we establish orderings of EGE distributions in terms of shape parameters.

Theorem 2.3. Let $X \sim EGE(\alpha_1, \beta, \lambda)$ and $Y \sim EGE(\alpha_2, \beta, \lambda)$, where $\alpha_1 \leq \alpha_2$. Then

- (i) $X \stackrel{lr}{\leq} Y$, $X \stackrel{hr}{\leq} Y$, $X \stackrel{st}{\leq} Y$ and $X \stackrel{mrl}{\leq} Y$.
(ii) $X \stackrel{dis}{\leq} Y$, for $\alpha_1 < 1$, $\beta < 0$.

Proof. Let $\beta \neq 0$. (i) From (1.2), we have

$$\frac{g(x)}{f(x)} = \frac{\alpha_2}{\alpha_1} \left[1 - (1 - \beta\lambda y)^{1/\beta} \right]^{\alpha_2 - \alpha_1},$$

where it is an increasing function of y . So $X \stackrel{lr}{\leq} Y$ and by Remark 2.2, other orderings of part (i) are obtained. (ii) For $\alpha_1 < 1$ and $\beta < 0$, EGE distribution has decreasing hazard rate (DHR), hence using the fact that if $X \stackrel{hr}{\leq} Y$ and X or Y is DHR, then $X \stackrel{dis}{\leq} Y$ (Shaked and Shantikumar (1994)), the result follows. Note that when $\beta = 0$, EGE distribution reduces to GE and orderings of part (i) and (ii) have been established by Abbasi Fard *et al.* (2011). \square

Theorem 2.4. Let $X \sim EGE(\alpha, \beta_1, \lambda)$ and $Y \sim EGE(\alpha, \beta_2, \lambda)$, where $\beta_1 \leq \beta_2$. Then

- (i) $Y \stackrel{dis}{\leq} X$ and $Y \stackrel{st}{\leq} X$.
(ii) $Y \stackrel{hr}{\leq} X$ and $Y \stackrel{mrl}{\leq} X$ for $\alpha > 1$, $\beta_2 > 1$.
(iii) $Y \stackrel{c}{\leq} X$ and $Y \stackrel{*}{\leq} X$.

Proof. Let $\beta \neq 0$, again the case $\beta = 0$ has been discussed by Abbasi Fard *et al.* (2011). (i) Using (1.2), we can write

$$F^{-1}(u) - G^{-1}(u) = \frac{1 - (1 - u^{1/\alpha})^{\beta_2}}{\beta_2\lambda} - \frac{1 - (1 - u^{1/\alpha})^{\beta_1}}{\beta_1\lambda}.$$

By differentiating with respect to u , it follows that it is an increasing function of u and so $Y \stackrel{dis}{\leq} X$. Also by Remark 2.2, $Y \stackrel{st}{\leq} X$. (ii) For $\alpha > 1$ and $\beta > 0$, EGE distribution has increasing hazard rate (IHR), hence using the fact that if $X \stackrel{dis}{\leq} Y$ and X or Y is IHR, then $X \stackrel{hr}{\leq} Y$ (Shaked and Shantikumar (1994)), the result follows. (iii) It is not so hard to show that $F^{-1}(G(y)) = \frac{1 - (1 - \beta_1\lambda y)^{\beta_2/\beta_1}}{\beta_2\lambda}$ is a convex function of y . So $Y \stackrel{c}{\leq} X$ and by Remark 2.2 $Y \stackrel{*}{\leq} X$. \square

3 Entropy Measures

Ever since Shannon (1948) has proposed a measure of uncertainty in a discrete distribution based on the Boltzmann entropy, there has been a great deal of interest in the measurement of uncertainty associated with a probability distribution. The Shannon entropy of a non-negative absolutely continuous random variable X with probability density function $f(x)$ is defined by

$$H(X) = - \int_0^{+\infty} f(x) \ln f(x) dx.$$

With his work, a new branch of mathematics with applications in different areas such as physics, probability and statistics, communication theory and economics was opened up.

Renyi (1961) defined the entropy of order γ as

$$H_\gamma(X) = - \frac{1}{\gamma-1} \ln \int_0^{+\infty} f^\gamma(x) dx, \quad \forall \gamma > 0 (\gamma \neq 1).$$

Rao *et al.* (2004) enumerated several drawbacks of $H(X)$. In response, Rao *et al.* (2004) and Wang *et al.* (2003) defined a new measure of uncertainty based on the survival function of a positive random variable instead of its density function and called it cumulative residual entropy (CRE). The survival function is $\bar{F}(x) = P(X > x) = 1 - F(x)$. The CRE is defined as follows

$$\mathcal{E}(X) = - \int_0^{+\infty} \bar{F}(x) \ln \bar{F}(x) dx.$$

They have obtained several properties of the CRE and provided some applications of it in reliability engineering and computer vision. Following this philosophy, Abbasnejad *et al.* (2010) introduced a measure of uncertainty, called the survival entropy. This measure is defined by

$$\mathcal{E}_\gamma(X) = - \frac{1}{\gamma-1} \ln \int_0^{+\infty} \bar{F}^\gamma(x) dx \quad \forall \gamma > 0 (\gamma \neq 1).$$

Two other measures of entropy, which are similar to CRE and SE, were introduced by Di Crescenzo and Longobardi (2009) and Abbasnejad (2011). These measures are obtained by replacing the survival function by cdf. Cumulative entropy (CE) is defined as

$$\mathcal{CE}(X) = - \int_0^{+\infty} F(x) \ln F(x) dx$$

and Failure entropy (FE) of order γ (FE) is as

$$\mathcal{F}\mathcal{E}_\gamma(X) = -\frac{1}{\gamma-1} \ln \int_0^{+\infty} F^\gamma(x) dx \quad \forall \gamma > 0 (\gamma \neq 1).$$

By some calculations, the closed form of entropy measures of EGE distribution are given as:

- Shannon entropy

$$H(X) = -\ln(\alpha\lambda) - (\alpha-1)\psi(\alpha) + (\beta-1)\psi(1) + (\alpha-\beta)\psi(\alpha+1),$$

where $\psi(\alpha) = \frac{d \ln \Gamma(\alpha)}{d\alpha}$.

- Renyi entropy

$$H_\gamma(X) = -\ln(\alpha\lambda) - \frac{1}{\gamma-1} [\ln \alpha + \ln B(\gamma(\alpha-1)+1, (1-\gamma)(\beta-1)+1)],$$

where $B(\gamma(\alpha-1)+1, (1-\gamma)(\beta-1)+1)$ is the beta function with parameters $\gamma(\alpha-1)+1$ and $(1-\gamma)(\beta-1)+1$.

- Cumulative residual entropy

$$\mathcal{E}(X) = \begin{cases} \frac{1}{\lambda} \sum_{i=1}^{\infty} \frac{1}{i} [B(\alpha i + 1, \beta) - B(\alpha(i+1) + 1, \beta)] & \text{if } \beta \neq 0 \\ \frac{1}{\lambda} \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} \frac{1}{i} \left[\frac{1}{\alpha i + j + 1} - \frac{1}{\alpha i + \alpha + j + 1} \right] & \text{if } \beta = 0. \end{cases}$$

- Survival entropy

$$\mathcal{E}_\gamma(X) = \begin{cases} \frac{1}{\gamma-1} \ln \lambda - \frac{1}{\gamma-1} \ln \sum_{i=0}^{\infty} (-1)^i \binom{\gamma}{i} B(\alpha i + 1, \beta) & \text{if } \beta \neq 0 \\ \frac{1}{\gamma-1} \ln \alpha \lambda - \frac{1}{\gamma-1} \ln \sum_{i=0}^{\infty} \frac{(-1)^i}{i} \binom{\gamma}{i} & \text{if } \beta = 0. \end{cases}$$

- Cumulative entropy

$$\mathcal{C}\mathcal{E}(X) = \frac{\alpha}{\lambda} \sum_{i=1}^{\infty} \frac{B(\alpha+1, \beta+i)}{i}.$$

- Failure entropy

$$\mathcal{F}\mathcal{E}_\gamma(X) = \begin{cases} \frac{1}{\gamma-1} \ln \lambda - \frac{1}{\gamma-1} \ln B(\gamma\alpha, \beta) & \text{if } \beta \neq 0 \\ \frac{1}{\gamma-1} \ln \lambda - \frac{1}{\gamma-1} \ln \sum_{i=0}^{\infty} \frac{(-1)^i}{i} \binom{\alpha\gamma}{i} & \text{if } \beta = 0. \end{cases}$$

Definition 3.1. The random variable X is said to be less than or equal to Y in the

- Shannon entropy ordering, denoted by $X \stackrel{Sh}{\leq} Y$, if and only if $H(X) \leq H(Y)$.
- Renyi entropy ordering, denoted by $X \stackrel{RE}{\leq} Y$, if $H_\gamma(X) \leq H_\gamma(Y)$.
- Cumulative residual entropy ordering, denoted by $X \stackrel{CRE}{\leq} Y$, if $\mathcal{E}(X) \leq \mathcal{E}(Y)$.
- Survival entropy ordering, denoted by $X \stackrel{SE}{\leq} Y$, if $\mathcal{SE}_\gamma(X) \geq \mathcal{SE}_\gamma(Y)$.
- Cumulative entropy ordering, denoted by $X \stackrel{CE}{\leq} Y$, if $\mathcal{CE}(X) \leq \mathcal{CE}(Y)$.
- Failure entropy ordering, denoted by $X \stackrel{FE}{\leq} Y$, if $\mathcal{FE}_\gamma(X) \geq \mathcal{FE}_\gamma(Y)$.

Theorem 3.2. Let $X \sim EGE(\alpha_1, \beta, \lambda)$ and $Y \sim EGE(\alpha_2, \beta, \lambda)$, where $\alpha_1 \leq \alpha_2$. Then

- (i) $X \stackrel{Sh}{\leq} Y$ and $X \stackrel{RE}{\leq} Y$, for $\alpha_1 < 1$, $\beta < 0$.
- (ii) $X \stackrel{CRE}{\leq} Y$ and $X \stackrel{CE}{\geq} Y$.
- (iii) $X \stackrel{SE}{\geq} (\leq) Y$, $\forall \gamma > 1$ ($0 < \gamma < 1$).
- (iv) $X \stackrel{FE}{\leq} (\geq) Y$, $\forall \gamma > 1$ ($0 < \gamma < 1$).

Proof. (i) Oja (1981) showed that $X \stackrel{dis}{\leq} Y$ results $X \stackrel{Sh}{\leq} Y$. Also if $X \stackrel{dis}{\leq} Y$, then $X \stackrel{RE}{\leq} Y$ (Abbasnejad (2008)). So the result is obtained by Theorem 2.3. (ii) CRE ordering is obtained by noting that if $X \stackrel{st}{\leq} Y$ then $X \stackrel{CRE}{\leq} Y$, (Zardasht (2015)) and using Theorem 2.3. By following the method of Zardasht (2015), it is not very hard to show that if $X \stackrel{st}{\leq} Y$ then $X \stackrel{CE}{\leq} Y$. Hence we have $X \stackrel{CE}{\leq} Y$ by Theorem 2.3. The proof of (iii) and (iv) is completed by Theorem 1 of Abbasnejad (2011b) and Theorem 5 of Abbasnejad *et al.* (2010). \square

Theorem 3.3. Let $X \sim EGE(\alpha, \beta_1, \lambda)$ and $Y \sim EGE(\alpha, \beta_2, \lambda)$, where $\beta_1 \leq \beta_2$. Then

- (i) $X \stackrel{Sh}{\geq} Y$ and $X \stackrel{RE}{\geq} Y$.
- (ii) $X \stackrel{CRE}{\geq} Y$ and $X \stackrel{CE}{\leq} Y$.
- (iii) $X \stackrel{SE}{\leq} (\geq) Y$, $\forall \gamma > 1$ ($0 < \gamma < 1$).
- (iv) $X \stackrel{FE}{\geq} (\leq) Y$, $\forall \gamma > 1$ ($0 < \gamma < 1$).

Proof. The proof is similar to Theorem 3.2. \square

References

- Abbasi Fard, A., Abbasnejad, M. and Mohammadi, D. (2010). Stochastic orderings of the generalized exponential distribution, *Proceeding of the 8th seminar probability and stochastic processes*.
- Abbasnejad M. (2011). Some Ordering Results Based on Failure Entropy, *Proceeding of the 8th seminar probability and stochastic processes*, 335–341.
- Abbasnejad, M. and Arghami, N.R. (2008). Some results on dynamic Renyi entropy, *Proceeding of the ninth Statistics Conference of Iran*, 1–11.
- Abbasnejad, M., Arghami, N.R. Morgenthaler S. and Mohtashami Borzadaran G.R. (2010) On the dynamic survival entropy, *Statistics and Probability letters*, **80**, 1962-1971.
- Baklizi, A. (2008). Likelihood and bayesian estimation of $Pr(X < Y)$ using lower record values from the generalized exponential distribution *Comput. Statist. Data Anal.*, **52**, 3468-3473.
- Barlow, R.E. and Proschan, F. (1975). *Statistical theory of reliability and life testing*, Holt, Rinehart, and Winston, New York, NY.
- Chen, D.G. and Lio, Y.L. (2010). Parameter estimations for generalized exponential distribution under progressive type-I interval censoring, *Comput. Statist. Data Anal.*, **54**, 1581-1591.
- Ebrahimi, N. and Pellerey, F. (1995). New partial ordering of survival functions based on the notion of uncertainty, *J. of Applied Probability*, **32**, 202–211.
- Ebrahimi, N., Maasoumi, E. and Soofi E.S. (1999). Ordering univariate distributions by entropy and variance, *J. of Econometrics*, **90**, 317–336.
- Faggioli, E. and Pellerey, F. (1993). New partial orderings and applications, *Naval Res. Logist.*, **40**, 829–842.
- Gupta, R.D. and Kundu, D. (1999). Generalized exponential distributions, *Austral. New Zealand J. Statist.*, **41**, 173-188.

- Gupta, R.D. and Kundu, D. (2001a). Generalized exponential distribution: Different method of estimations, *J. Statist. Comput. Simul.*, **69**, 315-338.
- Gupta, R.D. and Kundu, D. (2001b). Exponentiated exponential family: An alternative to gamma and Weibull distributions, *Biometrical J.*, **43**, 117-130.
- Gupta, R.D. and Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments, *J. Statist. Plan. Infer.*, **137**, 3537-3547.
- Kundu, D. and Gupta, R.D. (2008). Generalized exponential distribution: Bayesian estimation, *Comput. Statist. Data Anal.*, **52**, 1873-1883.
- Kundu, D. and Gupta, R.D. (2011). An extension of the generalized exponential distribution, *Statistical Methodology*, **8**, 485-496.
- Madi, M.T. and Raqab, M.Z. (2009). Bayesian inference for the generalized exponential distribution based on progressively censored data, *Commun. Statist. Theory Meth.*, **38**, 2016-2029.
- Nanda, A.K. and Paul, P. (2006). Some properties of past entropy and their applications, *Metrika*, **64**, 47-61.
- Raqab, M.Z. (2002). Inferences for generalized exponential distribution based on record statistics, *J. Statist. Plan. Infer.*, **104**, 339-350.
- Raqab, M.Z. and Madi, M.T. (2005). Bayesian inference for the generalized exponential distribution, *J. Statist. Comput. Simul.*, **75**, 841-852.
- Shaked M. and Shanthikumar J.G. (1994). *Stochastic Orders and Their Applications*, Academic Press, New York.
- Zardasht, V. (2015). A test for the increasing convex order based on the cumulative residual entropy, *J. of the Korean Statistical Society*, 1-11.



Estimation of stress-strength reliability based on generalized Rayleigh records

M. Abdi¹, A. Asgharzadeh²

¹Department of Statistics, Higher Education Complex of Bam, Bam, Iran

² Department of Statistics, University of Mazandaran, Babolsar, Iran

Abstract: This paper considers the estimation of the stress-strength reliability based on two-parameter generalized Rayleigh lower records. The maximum likelihood and Bayes estimators are derived. Bayesian credible interval and two bootstrap confidence intervals are also proposed. Finally, the analysis of a real data set has been presented for illustrative purposes.

Keywords Bayes estimator, Bootstrap confidence intervals, Generalized Rayleigh distribution, Maximum likelihood estimator.

Mathematics Subject Classification (2010): 62F10, 62F15, 62N05.

1 Introduction

The definition of record values was formulated by [Chandler \(1952\)](#). Let X_1, X_2, \dots be a sequence of independent and identically distributed (iid) continuous random variables with cdf $F(x)$ and pdf $f(x)$. If we define $T_1 = 1$, $U_1 = X_1$, and for $n \geq 2$,

$$T_n = \min\{j > T_{n-1} : X_j > X_{T_{n-1}}\}, \quad U_n = X_{T_n}.$$

Then the sequence $\{U_n\}(\{T_n\})$ is known as upper record statistics (upper record times). Similarly, the lower record times S_n and the lower record values L_n are defined as follows: $S_1 = 1$, $L_1 = X_1$, and for $n \geq 2$, $S_n = \min\{j > S_{n-1} : X_j < X_{S_{n-1}}\}$, $L_n = X_{S_n}$. For more details and applications of record values, see, for example, [Ahsanullah \(1995\)](#) and [Arnold et al. \(1998\)](#).

¹Speaker's Name :me.abdi@bam.ac.ir

The estimation of $R = P(Y < X)$ is very common in the statistical literatures. Suppose that X is the strength of a component which is subject to stress Y . The system fails if and only if at any time the applied stress is greater than the strength. Since R represents a relation between the stress and strength of a system, it is popularly known as the stress-strength parameter. Estimation of the stress-strength parameter has received considerable attention in the statistical literature. The problem of estimating R has been extensively studied for many statistical models based on complete samples. The monograph by [Kotz et al. \(2003\)](#) provided an excellent review of the development of this model till that time. Recently, some authors have studied the inferential procedures of R for some lifetime distributions based on record and censored samples. See, for example, [Asharzadeh et al. \(2011\)](#), [Saraoglua et al. \(2012\)](#), [Asharzadeh et al. \(2013\)](#) and [Asharzadeh et al. \(2014\)](#).

In this paper, we consider the estimation of the stress-strength parameter $R = P(X > Y)$, based on lower record values on both variables X and Y , when X and Y are independent generalized Rayleigh random variables. The generalized Rayleigh distribution (GRD) has the following distribution function and probability function;

$$F(x, \alpha, \lambda) = (1 - e^{-\lambda x^2})^\alpha, \quad x > 0, \quad \alpha > 0, \quad \lambda > 0, \quad (1.1)$$

$$f(x, \alpha, \lambda) = 2\alpha\lambda x e^{-\lambda x^2} (1 - e^{-\lambda x^2})^{\alpha-1}, \quad x > 0, \quad \alpha > 0, \quad \lambda > 0. \quad (1.2)$$

Here α and λ are the shape and scale parameters respectively. Note that the GRD is a particular member of the exponentiated Weibull distribution, originally proposed by [Mudholkar and Srivastava \(1993\)](#). For an excellent review of these distribution, the readers are referred to [Johnson et al. \(1995\)](#) and [Madi and Raqab \(2011\)](#). From now on, GRD with the shape parameter α and scale parameter λ will be denoted by $GR(\alpha, \lambda)$.

2 Estimation of R with same scale parameters

Let $X \sim GR(\alpha, \lambda)$ and $Y \sim GR(\beta, \lambda)$ with unknown shape parameters α , β and common scale parameter λ , where X and Y are independently distributed. The stress-strength parameter, R is

$$R = P[X > Y] = \frac{\alpha}{\alpha + \beta}. \quad (2.1)$$

Our interest is in estimating R based on lower record values on both variables.

2.1 Maximum likelihood estimation of R

Let $X_1 > X_2 > \cdots > X_n$ be the first n lower record values observed from $GR(\alpha, \lambda)$. Suppose also that $Y_1 > Y_2 > \cdots > Y_m$ are the m lower records observed from $GR(\beta, \lambda)$ independently from the first sample. The likelihood function of α , β and λ is given (see Arnold et al. (1998)) by

$$L(\alpha, \beta, \lambda) = \left[f(x_n, \alpha, \lambda) \prod_{i=1}^{n-1} \frac{f(x_i, \alpha, \lambda)}{F(x_i, \alpha, \lambda)} \right] \left[f(y_m, \alpha, \lambda) \prod_{j=1}^{m-1} \frac{f(y_j, \alpha, \lambda)}{F(y_j, \alpha, \lambda)} \right]. \quad (2.2)$$

Using (1.1) and (1.2), we obtain the likelihood function as

$$L(\alpha, \beta, \lambda) = 2^{n+m} \alpha^n \beta^m \lambda^{n+m} (1 - e^{-\lambda x_n^2})^\alpha (1 - e^{-\lambda y_m^2})^\beta e^{-\lambda(\sum_{i=1}^n x_i^2 + \sum_{j=1}^m y_j^2)} \\ \times e^{\sum_{i=1}^n \ln x_i + \sum_{j=1}^m \ln y_j} e^{-\sum_{i=1}^n \ln[1 - e^{-\lambda x_i^2}]} e^{-\sum_{j=1}^m \ln[1 - e^{-\lambda y_j^2}]}. \quad (2.3)$$

The log-likelihood function of α , β and λ is given by

$$\ln L(\alpha, \beta, \lambda) = (n + m) \ln 2 + n \ln \alpha + m \ln \beta + (n + m) \ln \lambda \\ + \alpha \ln(1 - e^{-\lambda x_n^2}) + \beta \ln(1 - e^{-\lambda y_m^2}) \\ + \sum_{i=1}^n \ln x_i + \sum_{j=1}^m \ln y_j - \lambda \left(\sum_{i=1}^n x_i^2 + \sum_{j=1}^m y_j^2 \right) \\ - \sum_{i=1}^n \ln[1 - e^{-\lambda x_i^2}] - \sum_{j=1}^m \ln[1 - e^{-\lambda y_j^2}]. \quad (2.4)$$

The MLEs of α , β and λ , say, $(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ can be obtained numerically as the solution of likelihood equations

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} + \ln(1 - e^{-\lambda x_n^2}) = 0, \quad (2.5)$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{m}{\beta} + \ln(1 - e^{-\lambda y_m^2}) = 0, \quad (2.6)$$

and

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n + m}{\lambda} + \frac{\alpha x_n^2 e^{-\lambda x_n^2}}{1 - e^{-\lambda x_n^2}} + \frac{\beta y_m^2 e^{-\lambda y_m^2}}{1 - e^{-\lambda y_m^2}}$$

$$-\left(\sum_{i=1}^n x_i^2 + \sum_{j=1}^m y_j^2\right) - \sum_{i=1}^n \frac{x_i^2 e^{-\lambda x_i^2}}{1 - e^{-\lambda x_i^2}} - \sum_{j=1}^m \frac{y_j^2 e^{-\lambda y_j^2}}{1 - e^{-\lambda y_j^2}} = 0. \quad (2.7)$$

Now, due to the invariance property of the ML estimators, the MLE of R is

$$\widehat{R} = \frac{\widehat{\alpha}}{\widehat{\alpha} + \widehat{\beta}}. \quad (2.8)$$

2.2 Bootstrap confidence intervals

In this section, two bootstrap methods are presented to construct the bootstrap confidence intervals of R . The first one is the percentile bootstrap confidence interval (boot-p) and the second one is the bootstrap-t confidence interval (Boot-t).

(i): Boot-p method

1. From the original two samples of lower record x_1, \dots, x_n and y_1, \dots, y_m , compute $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\lambda}$.
2. Using $\widehat{\alpha}$ and $\widehat{\lambda}$, generate a bootstrap sample $\{x_1^*, \dots, x_n^*\}$ and similarly using $\widehat{\beta}$ and $\widehat{\lambda}$, generate a bootstrap sample $\{y_1^*, \dots, y_m^*\}$. Based on the bootstrap samples $\{x_1^*, \dots, x_n^*\}$ and $\{y_1^*, \dots, y_m^*\}$, compute the bootstrap estimate of R using (2.12), say \widehat{R}^* .
3. Repeat 2, NBOOT times.
4. Let $h_1(x) = P(\widehat{R}^* \leq x)$ be the cumulative distribution function of \widehat{R}^* . Define $\widehat{R}_{Boot-p}(x) = h_1^{-1}(x)$ for a given x . The approximate $100(1 - \alpha)\%$ confidence interval of R is given by

$$\left(\widehat{R}_{Boot-p}^*\left(\frac{\alpha}{2}\right), \widehat{R}_{Boot-p}^*\left(1 - \frac{\alpha}{2}\right)\right)$$

(i): Boot-t Procedure:

1. From the original two samples of lower record x_1, \dots, x_n and y_1, \dots, y_m , compute $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\lambda}$ and \widehat{R} .
2. Same as in Boot-p method, first generate the bootstrap samples $\{x_1^*, \dots, x_n^*\}$, $\{y_1^*, \dots, y_m^*\}$ and then compute \widehat{R}^* , the bootstrap estimate of R . Also, compute the statistic

$$T^* = \frac{\sqrt{n}(\widehat{R}^* - \widehat{R})}{\sqrt{\widehat{V}(\widehat{R}^*)}}.$$

and its $Var(\widehat{R}^*)$ can be obtained using the Fisher information matrix.

3. Repeat Step 1 and 2, NBOOT times.

4. Let $h_2(x) = P(T^* \leq x)$ be the cumulative distribution function of T^* . For a given x define $\widehat{R}_{Boot-t}(x) = \widehat{R} + h_2^{-1}(x)\sqrt{\frac{Var(\widehat{R}^*)}{n}}$ The approximate $100(1 - \alpha)\%$ confidence interval of R is

$$\left(\widehat{R}^*_{Boot-p}\left(\frac{\alpha}{2}\right), \widehat{R}^*_{Boot-p}\left(1 - \frac{\alpha}{2}\right)\right).$$

2.3 Bayes estimation of R

It is assumed that α , β and λ have density functions $Gamma(a_1, b_1)$, $Gamma(a_2, b_2)$ and $Gamma(a_3, b_3)$, respectively. Moreover, it is assumed that α , β and λ are independent. Based on the above assumptions, we have the likelihood function of the observed data as

$$\begin{aligned} L(data|\alpha, \beta, \lambda) &= 2^{n+m} \alpha^n \beta^m \lambda^{n+m} (1 - e^{-\lambda x_n^2})^\alpha (1 - e^{-\lambda y_m^2})^\beta e^{-\lambda(\sum_{i=1}^n x_i^2 + \sum_{j=1}^m y_j^2)} \\ &\times e^{\sum_{i=1}^n \ln x_i + \sum_{j=1}^m \ln y_j} e^{-\sum_{i=1}^n \ln[1 - e^{-\lambda x_i^2}]} e^{-\sum_{j=1}^m \ln[1 - e^{-\lambda y_j^2}]} \end{aligned} \quad (2.9)$$

The joint density of the data, α , β and λ is

$$L(data, \alpha, \beta, \lambda) = L(data|\alpha, \beta, \lambda) \times \pi_1(\alpha) \times \pi_2(\beta) \times \pi_3(\lambda), \quad (2.10)$$

where $\pi_1(\alpha)$, $\pi_2(\beta)$ and $\pi_3(\lambda)$ are prior pdfs for α , β and λ , respectively. Therefore the joint posterior density of α , β and λ given the data is

$$\pi(\alpha, \beta, \lambda|data) = \frac{L(data|\alpha, \beta, \lambda)}{\int_0^\infty \int_0^\infty \int_0^\infty L(data|\alpha, \beta, \lambda) d\alpha d\beta d\lambda}, \quad (2.11)$$

Since (2.11) cannot be obtained analytically, we use the Gibbs sampling technique to compute the Bayes estimate of R and the corresponding credible interval of R . The posterior pdfs of α , β and λ are as follows:

$$\alpha|\beta, \lambda, data \sim Gamma\left(n + a_1, b_1 - \ln(1 - e^{-\lambda x_n^2})\right), \quad (2.12)$$

$$\beta|\alpha, \lambda, data \sim Gamma\left(m + a_2, b_2 - \ln(1 - e^{-\lambda y_m^2})\right), \quad (2.13)$$

$$f(\lambda|\alpha, \beta, data) \propto \lambda^{n+m+a_3-1} e^{-\lambda(b_3 + \sum_{i=1}^n x_i^2 + \sum_{j=1}^m y_j^2)} e^{\sum_{i=1}^n \ln x_i + \sum_{j=1}^m \ln y_j}$$

$$\times e^{-\sum_{i=1}^n \ln[1-e^{-\lambda x_i^2}]} e^{-\sum_{j=1}^m \ln[1-e^{-\lambda y_j^2}]} \quad (2.14)$$

The posterior pdf of λ is not known, so to generate samples from this distribution, we use the Metropolis method with normal proposal distribution. Therefore the algorithm of Gibbs sampling is as follows:

1. Start with an initial guess $(\alpha^{(0)}, \beta^{(0)}, \lambda^{(0)})$ and set $t = 1$.
2. Generate $\alpha^{(t)}$ from $Gamma\left(n + a_1, b_1 - \ln(1 - e^{-\lambda^{(t-1)} x_n^2})\right)$.
3. Generate $\beta^{(t)}$ from $Gamma\left(m + a_2, b_2 - \ln(1 - e^{-\lambda^{(t-1)} y_m^2})\right)$.
4. Using the Metropolis method, generate $\lambda^{(t)}$ from $f(\lambda^{(t-1)} | \alpha^{(t)}, \beta^{(t)}, data)$ with the $N(\lambda^{(t)}, 10)$ proposal distribution.
5. Compute $R^{(t)}$ from (2.1).
6. Set $t = t + 1$.
7. Repeat steps 2-6, T times.

Now the bayes estimator of R is the approximate posterior mean

$$\widehat{E}(R|data) = \frac{1}{T} \sum_{t=1}^T R^{(t)}. \quad (2.15)$$

Based on T and R values, one can construct a $100(1 - \gamma)\%$ bayesian credible interval as

$$\left(R_{[\frac{\gamma}{2}T]}, R_{[(1-\frac{\gamma}{2})T]} \right) \quad (2.16)$$

where $R_{[\frac{\gamma}{2}T]}$ and $R_{[(1-\frac{\gamma}{2})T]}$ are the $[\frac{\gamma}{2}T]$ -th smallest integer and the $[(1 - \frac{\gamma}{2})T]$ -th smallest integer of $\{R_t, t = 1, 2, \dots, T\}$, respectively.

2.4 Estimation of R with known scale parameter

In this section, we consider the estimation of R when the common scale parameter λ is known. Without loss of generality, we can assume that $\lambda = 1$. In this case, the MLE of R , say \widehat{R} can be obtained as

$$\widehat{R} = \frac{\widehat{\alpha}}{\widehat{\alpha} + \widehat{\beta}} \quad (2.17)$$

where

$$\widehat{\alpha} = \frac{-n}{\ln(1 - e^{-\lambda X_n^2})}, \quad \widehat{\beta} = \frac{-m}{\ln(1 - e^{-\lambda Y_m^2})}. \quad (2.18)$$

Therefore using the invariance properties of the maximum likelihood estimation, the MLE of R is

$$\widehat{R} = \frac{n \ln(1 - e^{-\lambda Y_n^2})}{n \ln(1 - e^{-\lambda Y_n^2}) + m \ln(1 - e^{-\lambda X_n^2})}. \quad (2.19)$$

It can be shown that

$$\widehat{R} = \frac{n \ln(1 - e^{-\lambda Y_n^2})}{n \ln(1 - e^{-\lambda Y_n^2}) + m \ln(1 - e^{-\lambda X_n^2})} = \frac{1}{1 + \frac{m\beta}{n\alpha} \frac{U}{V}} \quad (2.20)$$

where

$$U = -2\alpha \ln(1 - e^{-\lambda X_n^2}) \sim \chi^2(2n), \quad V = -2\beta \ln(1 - e^{-\lambda Y_m^2}) \sim \chi^2(2m).$$

Note that, by the independence of two random quantities U and V , we have $\frac{U/2n}{V/2m} \sim F_{(2n,2m)}$, so we can find that $\widehat{R} \stackrel{d}{=} \frac{1}{1 + \frac{\beta}{\alpha} F_{(2n,2m)}}$ and so

$$\frac{R}{1-R} \times \frac{1-\widehat{R}}{\widehat{R}} \sim F_{(2n,2m)}$$

Thus a $100(1 - \gamma)\%$ confidence interval for R is

$$\left[\frac{F_{\frac{\gamma}{2}, (2n, 2m)}}{F_{\frac{\gamma}{2}, (2n, 2m)} + (\frac{1}{R} - 1)}, \frac{F_{1-\frac{\gamma}{2}, (2n, 2m)}}{F_{1-\frac{\gamma}{2}, (2n, 2m)} + (\frac{1}{R} - 1)} \right].$$

For the Bayes estimation of R , consider the priors for α and β as the Gamma distributions $\alpha \sim \text{Gamma}(a_1, b_1)$ and $\beta \sim \text{Gamma}(a_2, b_2)$. It can be shown that the posterior pdf of R becomes

$$f_R(r) \propto \frac{r^{n+a_1-1}(1-r)^{m+a_1-1}}{[r(a_1 - \ln(1 - e^{-x_n^2})) + (1-r)(a_2 - \ln(1 - e^{-y_n^2}))]^{n+m+b_1+b_2}}. \quad (2.21)$$

Now, using the posterior pdf $f_R(r)$, Bayesian point and intervals estimates for R can be obtained.

3 Numerical Example

Here we consider a data analysis for two data sets reported by [Bennett and Filliben \(2000\)](#). They have reported minority electron mobility for p-type $Ga_{1-x}Al_xAs$ with seven different values of mole fraction. We use two data sets related to the mole fractions 0.25 and 0.30. These data are given as follows:

Data Set 1 (belongs to mole fraction 0.25): 3.051, 2.779, 2.604, 2.371, 2.214, 2.045, 1.715, 1.525, 1.296, 1.154, 1.016, 0.7948, 0.7007, 0.6292, 0.6175, 0.6449, 0.8881, 1.115, 1.397, 1.506, 1.528.

Data Set 2 (belongs to mole fraction 0.30): 2.658, 2.434, 2.288, 2.092, 1.959, 1.814, 1.530, 1.366, 1.165, 1.041, 0.9198, 0.7241, 0.6403, 0.576, 0.5647, 0.5873, 0.8013, 1.002, 1.250, 1.347, 1.368.

We used the Kolmogorov-Smirnov (K-S) tests for each data set to fit the generalized Rayleigh distribution. It is observed that for data sets 1 and 2, the K-S distances are 0.2453 and 0.2026 with the corresponding p-values 0.1395 and 0.3110, respectively. Therefore, it is clear that generalized Rayleigh distribution model fits well to both the data sets. For the above data, we observe that the first 15 values for both the data sets are the lower record values. The MLE and Bayes estimate of R become 0.5162 and 0.4816, respectively. To compute the Bayes estimate, since we do not have any prior information, we assumed that $a_1 = a_2 = a_3 = b_1 = b_2 = b_3 = 0.001$. The 95% Boot-p and Boot-t confidence intervals are obtained as (0.3366, 0.6680) and (0.2188, 0.7779), respectively. The credible interval is also, (0.3097, 0.6637).

References

- Ahsanullah, M. (1995). *Introduction to Record Statistics*, NOVA Science Publishers Inc., Huntington, New York.
- Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*, Wiley, New York.
- Asgharzadeh, A., Valiollahi, R. and Raqab, M. Z. (2011). Stress-strength reliability of Weibull distribution based on progressively censored samples, *SORT*, **35(2)**, 103–124.
- Asgharzadeh, A., Valiollahi, R. and Raqab, M. Z. (2013). Estimation of the stress-strength reliability for the generalized logistic distribution, *Statistical Methodology*, **15**, 73–94.
- Asgharzadeh, A., Valiollahi, R. and Raqab, M. Z. (2014). Estimation of $Pr(Y < X)$ for the two-parameter of generalized exponential records. To appear in *Communications in Statistics - Simulation and Computation*, <http://dx.doi.org/10.1080/03610918.2014.964046>

- Bennett, H. S. and Filliben, J. J. (2000). A systematic approach for multidimensional, closed form analytic modeling: minority electron mobilities in $Ga_{1-x}Al_xAs$ heterostructures, *Journal of Research of the National Institute of Standards and Technology*, 105 , 441–452.
- Chandler, K. N. (1952). The distribution and frequency of record values. *Journal of Royal Statistical Society*, B14, 220–228.
- Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, vol.1, 2nd Ed, New York, Wiley.
- Kotz, S., Lumelskii, Y. and Pensky, M. (2003). *The Stress-Strength Model and its Generalizations*, World Scientific Press, Singapore.
- Madi, M. T. and Raqab, M. Z. (2011). Statistical Inferences Based on Generalized Rayleigh Records. *6th Annual International Conference on Mathematics, Statistics and Education*. June 12–15, Athens, Greece.
- Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure data, *IEEE Transactions on Reliability*, 42, 299–302
- Saraoglua, B., Kinacia, I. and Kundu, D. (2012). On estimation of $R = P(Y < X)$ for exponential distribution under progressive type-II censoring. *Journal of Statistical Computation and Simulation* **85(2)**, 729–744.



Wavelet Shrinkage Estimator Of Regression Function Based on Maximum A Posterior

Mahmoud Afshari¹, Behzade Gholizadeh²

¹Department of Statistics, College of Science, Persian Gulf University ,
Bushehr, Iran

²Ph.d Student of Persian Gulf University

Abstract: Using wavelets is one of ways of estimating regression function. In this paper the wavelet shrinkage estimator of nonparametric regression function are considered. We applying the appropriate prior distributions and value of thresholding base on maximum a posterior to obtain the purpose estimator .

Keywords Nonparametric Regression, Wavelet Transformation, Shrinkage, Prior distribution, Maximum A Posterior, Thresholding.

Mathematics Subject Classification (2010): 62G08 65T60 62F15.

1 Int

The attention of the statistical community was attracted to wavelets when Daubechies (1992), discovery of compactly supported wavelet bases represents and established a connection between wavelets and signal processing. Since then, the wavelets have been proved useful in many statistical disciplines, notably in nonparametric statistics and time series analysis (see Afshari (2013), (2014), Doosti et al. (2008),and Antoniadis (2007). For deionising or shrinkage coefficients, one of the most important concepts in wavelets and deionising is using thresholds. So far, very methods have been presented to estimate the threshold function. It can be mentioned the Norm and Subband adaptive methods of recent works in this area, that proposed by Poornachandra (2008) respectively. Bayesian approaches to

²Speaker's Name : afshar@pgu.ac.ir

choose the shrinkage method are less ad hoc than some of the earlier proposals, and have been proved to be effective. For instance Kim and Akritas (2010) and Boubchir and Boashash (2013) suggested thresholding based on L-statistics and a nonparametric Bayesian estimator using Bessel K Form prior densities respectively. Some of bayesian paradigm and different bayesian wavelet regression models and methods for wavelet shrinkage can be fined in Shen and Zayed (2012) and Vidakovic (2009)and references therein. The purpose of this paper, is to provide a wavelet shrinkage estimator of regression function by using maximum a postorir(MAP) estimator. This article has three section as the following: Wavelet Shrinkage Estimator based on Maximum A Postorier, Numerical computation .

2 Wavelet Shrinkage Estimator based on Maximum A Postorier

Let the nested sequence of closed subspaces $\cdots V_{j-1} \subset V_j \subset V_{j+1} \subset \cdots$, $j \in \mathbb{Z}$, be a multiresolution approximation to $L^2(\mathbb{R})$. Define W_j , $j \in \mathbb{Z}$ to be orthogonal complement of V_j in V_{j+1} . The term wavelets are used to refer to set of basis function very special structure. The special of wavelet basis for function $f \in L^2(\mathbb{R})$ as scaling function ϕ and mother wavelet ψ such that $\{\phi(x - k)\}_{k \in \mathbb{Z}}$ forms an orthogonal basis for V_0 an $\{\psi(x - k)\}_{k \in \mathbb{Z}}$ forms an orthogonal basis for W_0 . Other wavelets in the basis are then generated by translation of the scaling function and dilations of the mother wavelet by using the following relationships:

$$\phi_{m,k}(x) = 2^{\frac{m}{2}} \phi(2^m x - k), \quad \text{and} \quad \psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^{\frac{j}{2}} x - k).$$

Two versions of the wavelet transformation can be distinguished: the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT). The difference between them lies in the set of scales and positions at which each transform operates. Let $\{X_n, n \geq 1\}$ be a sequence of random variables on the probability space $(\Omega, \mathfrak{R}, P)$. Suppose that X_i has a bounded and compactly supported marginal density $f(x)$, with respect to Lebesgue measure, which does not depend on i . We estimate this density from n observations X_i , $i = 1, \cdots, n$. For any function $f \in L^2(\mathbb{R})$, we can write a formal

expansion (see Daubechies (1992):

$$f = \sum_{k \in \mathbb{Z}} \alpha_{m,k} \phi_{m,k} + \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \delta_{j,k} \psi_{j,k} = P_m f + \sum_{j=m}^{\infty} D_j f, \quad (2.1)$$

Where the wavelet coefficients are given by the integrals

$$\alpha_{m,k} = \int f(x) \phi_{m,k}(x) dx, \quad \delta_{j,k} = \int f(x) \psi_{j,k} dx.$$

Let Y_1, \dots, Y_n are observed data from model,

$$Y_i = f(X_i) + \epsilon_i, \quad (2.2)$$

where the $\{\epsilon_i\}$ is some noise with density $N(0, 1)$ and $\{X_i\}$ is some points from domain of f which can be assumed by-the-way or regular. Typically n is an integer power of 2. The corresponding wavelet coefficient for $f \in L^2(\mathbb{R})$ will be written as $\theta_{j,k} = \langle f, \psi_{j,k} \rangle$. The indices $\{j, k\}$ are the standard notations for the level and shift of position in the multiresolution table of wavelet coefficients. Next, to simplify notation, indices are dropped and refer to d as an arbitrary wavelet coefficient from the detail spaces. If \mathbf{W} represent the discrete wavelet transform matrix, then multiplication of the equation (2.2) by orthogonal matrix \mathbf{W} yields:

$$\mathbf{d} = \mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{f} + \mathbf{W}\boldsymbol{\epsilon} = \boldsymbol{\theta} + \boldsymbol{\eta}.$$

The choice of the threshold is a very delicate and important statistical problem. On the one hand, a big threshold cuts too many coefficients and will result in an oversmoothing estimator. But on the other hand, a small threshold does not remove noise well and will produce a wiggly, undersmoothing estimator and increases the variance of the smoother. Theoretical considerations yield the following value of the threshold: $\lambda = \sqrt{2\sigma^2 \log n}$, where n is the length of the input vector and σ^2 is the variance of the noise. One of the popular methods in determining threshold is bayesian paradigm. Consider the following model:

$$d|\theta \sim N(\theta, \sigma^2), \quad \theta \sim \pi(\theta), \quad (2.3)$$

where σ^2 is known and in practice estimate by empirical approaches. We seek *maximum a posteriori*(MAP) estimator, defined as the posterior mode. Notice that the posterior is proportional to the

joint distribution of d and θ . Therefore, the MAP estimator maximizes $l(\theta|d)\pi(\theta)$ as well. But for some priors the MAP estimator can not estimated in closed form. For example, if θ has Cauchy prior distribution, $\pi(\theta) = C(\tau, 1)$, then joint distribution of d, θ is the following:

$$f(d, \theta) = \frac{1}{\sqrt{2\pi^3\sigma^2}} \frac{\exp\left\{-\frac{1}{2\sigma^2}(d-\theta)^2\right\}}{1+(\theta-\tau)^2}.$$

It is clear that the posterior distribution of θ hasn't closed form so, the calculation of MAP estimator is difficult. One way to surmount this difficulties, is approximate prior distribution by conjugate family of distributions.

Theorem2.1: Let d be a transformed data from model such that, $d|\theta, \sigma^2 \sim N(\theta, \sigma^2)$, $\theta \sim \pi(\theta) = \sum_{j=1}^N w_j N(\eta_j, \tau_j^2)$ and modified posterior distribution of θ , considered as,

$$\pi(\theta|d) = w_0\delta_0 + \sum_{j=1}^N w'_j(d) N\left(\frac{\eta_j\sigma^2 + \tau_j^2 d}{\sigma^2 + \tau_j^2}, \frac{\sigma^2\tau_j^2}{\sigma^2 + \tau_j^2}\right),$$

where, $w_0 + \sum_{j=1}^N w_j = 1$ and δ_0 is degenerate distribution in zero. Then a threshold for θ can be considered as the following:

$$\hat{\theta} = \begin{cases} E^\pi(\theta|d), & f(d) \geq \frac{C_{II}w_0}{(1-w_0)C_I}, \\ 0, & o.w. \end{cases} \quad (2.4)$$

where $f(d)$ is marginal distribution of d and C_I and C_{II} are losses under H_0 and H_1 assumption respectively.

Proof: According to the bayesian approach in hypothesis testing, the thresholding estimator of θ can be write as the following:

$$\hat{\theta} = \hat{\theta}_\pi(d) \times \left(p(\theta \in \Theta_0|d) < \frac{C_I}{C_I + C_{II}} \right),$$

Since, $p(\theta \in \Theta_0|d) = \frac{w_0\delta_0}{w_0\delta_0 + (1-w_0)f(d)} < \frac{C_I}{C_I + C_{II}} \Leftrightarrow f(d) > \frac{C_{II}w_0}{(1-w_0)C_I}$, where,

$$\begin{aligned} f(d) &= \int_{-\infty}^{+\infty} f(d|\theta)\pi(\theta)d\theta = \sum_{j=1}^N w_j \int_{-\infty}^{+\infty} f(d|\theta)\pi_j(\theta)d\theta \\ &= \sum_{j=1}^N w_j \int_{-\infty}^{+\infty} \frac{1}{\sqrt{4\pi^2\tau_j^2\sigma^2}} \exp\left\{-\frac{1}{\sigma^2}(\theta-d)^2 - \frac{1}{\tau_j^2}(\eta_j-\theta)^2\right\} d\theta \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^N w_j \int_{-\infty}^{+\infty} \frac{1}{\sqrt{4\pi^2\tau_j^2\sigma^2}} \exp \left\{ -\frac{\tau_j^2 + \sigma^2}{2\tau_j^2\sigma^2} \left(\theta^2 - 2\theta \frac{\tau_j^2 d + \sigma^2 \eta_j}{\tau_j^2 + \sigma^2} \right) \right. \\
&\quad \left. - \left(\frac{d^2}{2\sigma^2} + \frac{\eta_j^2}{2\tau_j^2} \right) \right\} d\theta \\
&= \sum_{j=1}^N w_j \int_{-\infty}^{+\infty} \frac{1}{\sqrt{4\pi^2\tau_j^2\sigma^2}} \exp \left\{ -\frac{\tau_j^2 + \sigma^2}{2\tau_j^2\sigma^2} \left(\theta^2 - 2\theta \frac{\tau_j^2 d + \sigma^2 \eta_j}{\tau_j^2 + \sigma^2} \right)^2 \right\} \\
&\quad \times \exp \left\{ \frac{\tau_j^2 + \sigma^2}{2\tau_j^2\sigma^2} \left(\frac{\tau_j^2 d + \sigma^2 \eta_j}{\tau_j^2 + \sigma^2} \right)^2 - \left(\frac{d^2}{2\sigma^2} + \frac{\eta_j^2}{2\tau_j^2} \right) \right\} d\theta \\
&= \sum_{j=1}^N w_j \int_{-\infty}^{+\infty} \frac{1}{\sqrt{4\pi^2\tau_j^2\sigma^2}} \exp \left\{ -\frac{\tau_j^2 + \sigma^2}{2\tau_j^2\sigma^2} \left(\theta - \frac{\tau_j^2 d + \sigma^2 \eta_j}{\tau_j^2 + \sigma^2} \right)^2 \right. \\
&\quad \left. - \frac{1}{2(\tau_j^2 + \sigma^2)} (d - \eta_j)^2 \right\} d\theta \\
&= \sum_{j=1}^N w_j \frac{1}{\sqrt{2\pi(\tau_j^2 + \sigma^2)}} \exp \left\{ -\frac{1}{\tau_j^2 + \sigma^2} (d - \eta_j)^2 \right\}.
\end{aligned}$$

therefore we can write the following thresholding estimator,

$$\hat{\theta} = \begin{cases} E^\pi(\theta|d), & f(d) \geq \frac{C_{II}w_0}{(1-w_0)C_I}, \\ 0, & o.w. \end{cases}$$

Remark2.1: If the variance of data was unknown, the threshold would be equal (2.4), but we using the estimation of variances instead of τ_j^2 -s in marginal distribution $f(d)$.

Example: Let $d|\theta \sim N(\theta, 1)$ and $\theta \sim N(\eta, \tau^2)$,

then, $f(d|\theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(\theta - d)^2 \right\} = \frac{\exp \left\{ -\frac{d^2}{2} \right\}}{\sqrt{2\pi}} \exp \{d\theta - \theta^2\} = h(d) \exp \{d\theta - \psi(\theta)\}$, such that $\psi(\theta) = \theta^2$ and $h(d) = \frac{1}{2\pi} \exp \left\{ -\frac{d^2}{2} \right\}$. Also,

$$\pi(\theta|\eta, \tau^2) = \frac{\exp \left\{ -\frac{\eta^2}{2\tau^2} \right\}}{\sqrt{2\pi\tau^2}} \exp \left\{ \theta \frac{\eta}{\tau^2} - \frac{1}{2\tau^2} \psi(\theta) \right\} = K(\mu, \lambda) \exp \{ \mu\theta - \lambda\psi(\theta) \},$$

where $\mu = \frac{\eta}{\tau^2}$, $\lambda = \frac{1}{2\tau^2}$ and $K(\mu, \lambda) = \sqrt{\frac{\lambda}{\pi}} \exp \left\{ -\frac{\mu^2}{4\lambda} \right\}$. So

$$\pi(\theta|d) = \sqrt{\frac{\tau^2 + 1}{2\pi\tau^2}} \exp \left\{ \theta \left(\frac{\eta + \tau^2 d}{2\tau^2} \right) \right\}$$

$$\begin{aligned}
& -\frac{\tau^2 + 1}{2\tau^2}\theta^2 \Big\} \exp \left\{ -\frac{\tau^2 + 1}{2\tau^2} \left(\frac{\eta - \tau^2 d}{\tau^2 + 1} \right) \right\} \\
& = K(\mu + d, \lambda + 1) \exp \{(\mu + d)\theta - (\lambda + 1)\psi(\theta)\}.
\end{aligned}$$

in particular case, suppose that $d|\theta \sim N(\theta, 1)$ and $\theta \sim \pi(\theta) = \sum_{j=1}^N w_j N(\eta_j, \tau_j^2)$. Then the posterior distribution of θ is equal to, $\pi(\theta|d) = \sum_{j=1}^N w'_j(d) N\left(\frac{\eta_j + \tau_j^2 d}{\tau_j^2 + 1}, \frac{\tau_j^2}{\tau_j^2 + 1}\right)$ where,

$$w'_j(d) = w_j \sqrt{\frac{\lambda_i}{\lambda_i + 1}} \exp \left\{ -\frac{\mu_i^2}{4\lambda_i} + \frac{(\mu_i + d)^2}{4(\lambda_i + 1)} \right\}$$

3 Numerical Computation

In this section the comparing is between the proposed method and three commonly-used shrinkage strategies. Hard and Soft thresholding with the universal rule, (Donaho (1995)), Hard and Soft thresholding with the BayesThresh rule, (Abramovich et al.(1998) and Robert et al. (2013)). To assess performance, we calculated the average mean squared error(AMSE) from the 100 simulations. We use two "Blocks" test functions, that are continuous functions and that have eight and fourteen jumps. Each experiment consist $N = 256, 512$ and 1024 observations from the test functions with the white noise is added to the sample data of them. As can be seen in Table 1 , by increasing the sample size,

Table 1: Average mean square error for three rule(universal, sure and heretical bayesian method) and our new proposed method. In the following f_i is the true signal and \hat{f}_i^j is the estimate of the function from simulation j

$AMSE = \frac{1}{100} \sum_{j=1}^{100} \sum_{i=1}^n \frac{(f(x_i) - \hat{f}(x_i^j))^2}{n}$					
N	Sure	Universal HardThresh	Universal SoftThresh	Hard BayesThresh	New Method
256	0.1158	0.5252	0.1939	0.1149	0.1789
512	0.0886	0.3365	0.1327	0.0823	0.0910
1024	0.0645	0.2061	0.0759	0.0544	0.0711

the accuracy of the proposed method, increases whereas this trend is not established in other ways

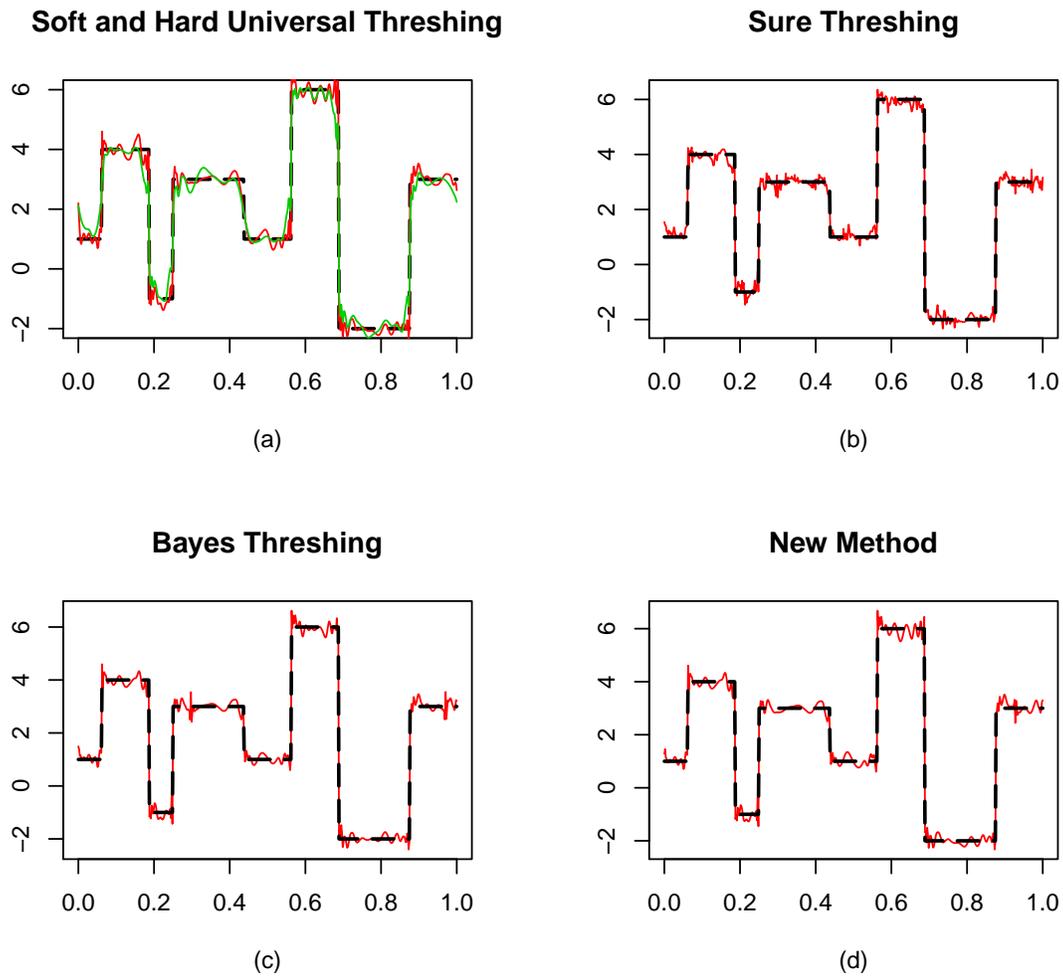


Figure 1: The panel (a) in Figure 1, display hard and soft universal wavelet estimator (red and green respectively) and the panel (b),(c),(d) display sure, heretical bayesian and proposed estimators respectively. The denoted line representing the true function and the number of samples is 512.

with the convergence speed of the proposed method. Also, the error of proposed method, in the worst case, is less than the global threshold method.

4 Conclusion

The main purpose of this paper was providing a new MAP wavelet shrinkage estimation of noisy signals. The proposed method applied for de-noising and estimating of parameters in simulation. According to the Table 1, it was shown that, the new proposed method has better AMSE with respect to other methods in most situations. The convergence ratio of this estimator is better than other methods. According to the figure1, nonparametric curve estimation by new thresholding wavelet is better than another methods. It was clear that the target estimator better than others.

Acknowledgement

The support of Research Committee of Persian Gulf University is greatly acknowledged.

References

- Abramovich, F., Theofanis, S., and Bernard, W.S. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60.4, 725-749.
- Afshari, M. (2013). Wavelet Density Estimation of Censoring Data and Evaluate of Mean Integral Square Error with convergence ratio and Empirical Distribution of Given Estimator. *Journal of Data Analysis and Information Processing*. 2,1-5.
- Afshari, M. (2014). A fast wavelet algorithm for analyzing of signal processing and empirical distribution of Wavelet coefficients with numerical example and simulation. *Communication of statistics-Theory and methods*. 42,4156-4169.
- Antoniadis, A. (2007). Wavelet methods in statistics: Some recent developments and their applications. *Statistics Surveys*, 1, 16-55.

- Boubchir, L., and Boashash, B. (2013). Wavelet denoising based on the MAP estimation using the BKF prior with application to images and EEG signals. *Signal Processing, IEEE Transactions on*, 61(8), 1880-1894.
- Daubechies, I. (1992). *Ten lectures on wavelets* (Vol. 61, pp. 198-202). Philadelphia: Society for industrial and applied mathematics.
- Donoho, D.L. (1995). De-noising by soft-thresholding. *Information Theory, IEEE Transactions on* 41.3,613-627.
- Doosti, H., Afshari, M., and Niroomand, H.A. (2008). Wavelets for nonparametric stochastic regression with mixing stochastic process. *Communications in Statistics Theory and Methods*. 37(3),373-385.
- Kim, M. H., and Akritas, M. G. (2010). Order thresholding. *The annals of statistics*, 38(4), 2314-2350.
- Poornachandra, S. (2008). Wavelet-based denoising using subband dependent threshold for ECG signals. *Digital Signal Processing*, 18(1), 49-55.
- Robert, C., and Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.
- Shen, X., and Zayed, A. I. (2012). *Multiscale Signal Analysis and Modeling*. Springer Science & Business Media.
- Vidakovic, B. (2009). *Statistical modeling by wavelets*. Wiley Series in Probability and Statistics, John Wiley & Sons, New York. Vol. 503.



On the Entropy Rate of Stationary Markov Chains

Mansour Aghababaei Jazi ¹, Hasan Zarei²

^{1,2}Department of Statistics, University of Sistan and Baluchestan, Zahedan, Iran

Abstract: In this paper we consider the entropy rate of (stationary) Markov chains and integer valued autoregressive processes as a special case. We study the rate for some orders and extend the result to general order.

Keywords Convolution, Entropy, Markov chain.

Mathematics Subject Classification (2010): 94A15, 60J10.

1 Introduction

A stochastic process X_1, X_2, \dots is an indexed sequence of random variables and is characterized by the joint probability distribution function (pdf)

$$\begin{aligned} Pr (X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= \\ &= \prod_{k=1}^n Pr(X_k = x_k | X_1 = x_1, \dots, X_{k-1} = x_{k-1}), \quad n = 1, 2, \dots \end{aligned} \quad (1.1)$$

Definition 1. X_1, X_2, \dots is said to be a stationary process, if the joint pdf (1) of any subset of the sequence of random variables is invariant with respect to shifts in the time index, i.e.,

$$Pr(X_1 = x_1, \dots, X_n = x_n) = Pr(X_{h+1} = x_1, \dots, X_{h+n} = x_n),$$

for every shift h and for all x_1, x_2, \dots, x_n .

Definition 2. X_1, X_2, \dots are said to be a Markov chain of k th order, if every random variable (r.v.) X_n depends on $X_{n-1}, X_{n-2}, \dots, X_{n-k}$ and is conditionally independent of all the other preceding r.v.'s $X_1, X_2, \dots, X_{n-k-1}$, i.e., if the transition probabilities can be simplified as

¹ aghababaei@math.usb.ac.ir

$$\begin{aligned} Pr(X_n = x_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1}) &= \\ &= Pr(X_n = x_n | X_{n-k} = x_{n-k} = x_{n-k}, \dots, X_{n-1} = x_{n-1}), \end{aligned}$$

for all x_1, x_2, \dots, x_n and $n = k + 1, k + 2, \dots$

Definition 3. If X_1, X_2, \dots are Markov chain of first order and the transition probabilities, $P_{ij}(n) \equiv Pr(X_n = j | X_{n-1} = i)$, do not depend on n , i.e., $P_{ij}(n) = Pr(X_2 = j | X_1 = i) \equiv P_{ij}$, then the chain is said to be time invariant (or homogeneous).

Also, the Markov chain of second order X_1, X_2, \dots is called time invariant, if $P_{hij}(n) \equiv Pr(X_n = j | X_{n-1} = i, X_{n-2} = h) = Pr(X_3 = j | X_2 = i, X_1 = h) \equiv P_{hij}$, $\forall n = 2, 3, \dots$. In this paper, Markov chains are assumed to be time invariant.

The marginal pdf of the Markov chain of first order X_1, X_2, \dots at time n , i.e., $\pi_i^{(n)} \equiv Pr(X_n = i)$ is characterized by

$$\begin{aligned} \pi_j^{(n)} &= \sum_i Pr(X_n = j, X_{n-1} = i) \\ &= \sum_i Pr(X_{n-1} = i) \cdot Pr(X_n = j | X_{n-1} = i) \\ &= \sum_i Pr(X_{n-1} = i) P_{ij}. \end{aligned}$$

Definition 4. If the pdf $\pi_i^{(n)}$ does not depend on n , i.e.

$$\pi_i^{(n)} = \pi_i^{(1)} \equiv \pi_i, \quad \forall n = 2, 3, \dots,$$

then the pdf is said to be the stationary pdf of the chain. In this case, the stationary pdf π_i is the solution to the equation system

$$\pi_j = \sum_i \pi_i P_{ij}$$

so that $\sum_j \pi_j = 1$.

Also, the marginal pdf of a Markov chain of second order X_1, X_2, \dots given by

$$\pi_j^{(n)} = \sum_i \sum_h Pr(X_{n-1} = i, X_{n-2} = h) P_{hij}$$

is called stationary, if $\pi_j^{(n)} = \pi_j$, for all $n = 1, 2, \dots$, and can be obtained by solving

$$\pi_j = \sum_i \sum_h \pi_h P_{hi} P_{hij}$$

so that $\sum_j \pi_j = 1$ (see, e.g., Resnick (2002) for more details).

In the next section, we recall the entropy rate of stochastic processes and stationary Markov Chains in two points of view and the estimation method of the rate. In Section 3, we consider the entropy rate of stationary integer valued autoregressive processes.

2 Entropy Rate of Stochastic Processes

The entropy of r.v. X with pdf $f(\cdot)$ has been defined to be $H(X) = -E(\log f(X))$. It implies that the entropy of r.v.'s X_1, X_2, \dots, X_n with joint pdf (1) is given by the following chain rule

$$H(X_1, X_2, \dots, X_n) = \sum_{k=1}^n H(X_k | X_1, X_2, \dots, X_{k-1}). \quad (2.1)$$

Definition 5. The entropy rate of a stochastic process X_1, X_2, \dots is defined by

$$H(\chi) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n),$$

when the limit exists.

The entropy rate has also been defined by

$$H'(\chi) = \lim_{n \rightarrow \infty} H(X_n | X_1, X_2, \dots, X_{n-1}),$$

when the limit exists. This rate is actually the entropy of the most recent output given all past outputs. Obviously, $H(\chi) = H'(\chi) = H(X_1)$ for a sequence of iid r.v.'s. The first equality also holds for stationary stochastic processes (Cover and Thomas (2006)).

Theorem 1. The entropy rate of stationary stochastic process X_1, X_2, \dots is given by

$$H(\chi) = \lim_{n \rightarrow \infty} H(X_n | X_1, X_2, \dots, X_{n-1}),$$

i.e., $H(\chi) = H'(\chi)$.

Proof. Let X_1, X_2, \dots be an stationary stochastic process. Then,

$$\begin{aligned} H(X_{n+1}|X_1, X_2, \dots, X_n) &\leq H(X_{n+1}|X_2, X_3, \dots, X_n) \\ &= H(X_n|X_1, X_2, \dots, X_{n-1}), \end{aligned}$$

where the inequality follows from the fact that conditioning reduces entropy and the equality follows from the stationarity of the process. Hence, $H(X_n|X_1, X_2, \dots, X_{n-1})$ is a decreasing sequence in n and, consequently,

$$H'(\chi) = \lim_{n \rightarrow \infty} H(X_n|X_1, X_2, \dots, X_{n-1})$$

exists. Also, from the chain rule (2), we obtain

$$H(\chi) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(X_i|X_1, X_2, \dots, X_{i-1})$$

where, by Cesaro mean, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(X_i|X_1, X_2, \dots, X_{i-1}) &= \lim_{n \rightarrow \infty} H(X_n|X_1, X_2, \dots, X_{n-1}) \\ &= H'(\chi). \end{aligned}$$

The above equalities implies that $H(\chi) = H'(\chi)$ for stationary stochastic processes.

2.1 Entropy Rate of Stationary Markov Chains

Here, we study the entropy rate of a stationary Markov chains. Let X_1, X_2, \dots be a stationary Markov chain of first order. Then the entropy rate is given by

$$\begin{aligned} H_1(\chi) &= \lim_{n \rightarrow \infty} H(X_n|X_1, X_2, \dots, X_{n-1}) \\ &= \lim_{n \rightarrow \infty} H(X_n|X_{n-1}) \\ &= H(X_2|X_1) \\ &= - \sum_i \sum_j \pi_i P_{ij} \log P_{ij} \\ &= \sum_i \pi_i \left(- \sum_j P_{ij} \log P_{ij} \right), \end{aligned} \tag{2.2}$$

i.e., a weighted sum of the entropy values for each state.

Also, the entropy rate of the stationary Markov chain of second order is given by

$$\begin{aligned} H_2(\chi) &= H(X_3|X_2, X_1) \\ &= \sum_h \pi_h \left(\sum_i P_{hi} \left(- \sum_j P_{hij} \log P_{hij} \right) \right) \end{aligned} \quad (2.3)$$

By stationary and entropy property, the entropy rate of the stationary Markov chain of second order is less than the first order, i.e.,

$$\begin{aligned} H_2(\chi) &= H(X_3|X_2, X_1) \\ &\leq H(X_3|X_2) \\ &= H(X_2|X_1) \\ &= H_1(\chi) \end{aligned}$$

Remark 1. Generally, it can be shown that the entropy rate of stationary Markov chain is non-increasing as the order is increasing, i.e,

$$\cdots \leq H_k(\chi) \leq H_{k-1}(\chi) \leq \cdots \leq H_2(\chi) \leq H_1(\chi) \leq H(X_1),$$

wherein

$$\begin{aligned} H_k(\chi) &= \lim_{n \rightarrow \infty} H(X_n|X_1, X_2, \dots, X_{n-1}) \\ &= \lim_{n \rightarrow \infty} H(X_n|X_{n-k}, X_{n-k+1}, \dots, X_{n-1}) \\ &= H(X_{k+1}|X_1, X_2, \dots, X_k), \quad k = 1, 2, \dots \end{aligned}$$

is the entropy rate of the stationary Markov chain of order k . This implies that

$$\begin{aligned} H_\infty(\chi) &\equiv \lim_{k \rightarrow \infty} H_k(\chi) \\ &= H'(\chi) \end{aligned}$$

exist (see, e.g., Cover and Thomas (2006) for more details).

2.2 Estimation of Entropy Rate

Let X_1, X_2, \dots, X_n be an observed Markov chain of first order and set

$$N_n(i, j) = \sum_{m=2}^n 1_{(X_{m-1}=i, X_m=j)}$$

and

$$N_n(i) = \sum_{m=1}^n 1_{(X_m=i)}$$

where $1_{(x=y)} = 1$, if $x = y$ otherwise $1_{(x=y)} = 0$. It can be shown (see, e.g., Sadek(2003)) that

$$\hat{P}_{ij} = \frac{N_n(i, j)}{N_n(i)}$$

are the MLE's of transition probabilities P_{ij} . Also, π_i is estimated by

$$\hat{\pi}_i = \frac{N_n(i)}{n}$$

Remark 2. From the law of large numbers and central limit theorem, it has been shown that the above estimators are strongly consistent and asymptotically normal, when n tends to infinity (see, e.g., Dacunha-Castelle and Duflo (1994) for more details).

Furthermore, the entropy rate of the Markov chain of first order (3) can be estimated by

$$\hat{H}_1(\chi) = - \sum_i \sum_j \hat{\pi}_i \hat{P}_{ij} \log \hat{P}_{ij}$$

Generally, the entropy rate estimation of Markov chains with more orders can be proposed by a similar argument.

3 Entropy Rate of INAR(k) Processes

Steutel and van Harn (1979) introduced the binomial thinning operator \circ . Let X be a nonnegative integer valued r.v. with pdf $\pi_i = Pr(X = i)$, $i = 0, 1, \dots$

Definition 6. The binomial thinning operator \circ is defined by

$$\alpha \circ X = \sum_{i=1}^X B_i(\alpha),$$

where counting series $B_i(\alpha)$ is a sequence of iid binary r.v.'s with $Pr(B_i(\alpha) = 1) = 1 - Pr(B_i(\alpha) = 0) = \alpha$ and $\alpha \in [0, 1]$.

The integer valued first order autoregressive (INAR(1)) process was independently introduced by Makenzie(1985) and Al-Osh and Alzaid(1987) based on the binomial thinning \circ operator.

Definition 7. Integer valued stochastic process X_1, X_2, \dots follows INAR(1) model, if

$$X_n = \alpha \circ X_{n-1} + \varepsilon_n, \quad n = 2, 3, \dots,$$

where ε_n 's known as innovations are iid nonnegative integer valued r.v.'s and independent of all the binary counting series and X_{n-1} .

Remark 3. If X_n is the population at time n then $\alpha \circ X_{n-1}$ and ε_n can be respectively interpreted as the number of survivors and the number of immigrants at time n from the previous period. It has been shown that an stationary INAR(1) process has a Poisson marginal pdf (with mean λ) if and only if the innovations also follow a Poisson pdf (with mean $\lambda(1 - \alpha)$). Also, the transition probabilities can be obtained as (5) in the following theorem.

Theorem 2. Let X_1, X_2, \dots follow a INAR(1) process with innovations ε_n 's following stationary pdf $q_k \equiv Pr(\varepsilon_n = k)$, $k = 0, 1, \dots$. Then, the process forms a stationary Markov chain with transition probabilities

$$\begin{aligned} P_{ij} &\equiv P(X_n = j | X_{n-1} = i) \\ &= \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1 - \alpha)^{i-k} q_{j-k}, \quad i, j = 0, 1, \dots, \end{aligned} \quad (3.1)$$

Remark 4. Clearly, P_{ij} as given in (5) is the convolution pdf of $\text{Bin}(i, \alpha)$ and q_k and giving the probability of going from state i to state j in a single step and can be set in (3) to yield the entropy rate of INAR(1) with stationary marginal pdf $\pi_i \equiv Pr(X_n = i)$.

Also, let X_1, X_2, \dots follow integer valued second order autoregressive (INAR(2)) model

$$X_n = \alpha \circ X_{n-1} + \beta \circ X_{n-2} + \varepsilon_n, \quad n = 3, 4, \dots,$$

where innovations ε_n 's are iid nonnegative integer valued r.v.'s and independent of all the binary counting series and X_{n-1} and X_{n-2} . If innovations ε_n 's follow stationary pdf $q_k \equiv Pr(\varepsilon_n = k)$,

$k = 0, 1, \dots$. Then, the process forms a stationary Markov chain of second order with transition probabilities

$$\begin{aligned} P_{hij} &\equiv P(X_n = j | X_{n-1} = i, X_{n-2} = h) \\ &= \sum_{k=0}^{\min(i,j)} \sum_{l=0}^{\min(j-k,h)} \binom{i}{k} \binom{h}{l} \alpha^k (1-\alpha)^{i-k} \beta^l (1-\beta)^{h-l} q_{j-k-l}, \\ h, i, j &= 0, 1, \dots, \end{aligned} \quad (3.2)$$

i.e., P_{hij} is the convolution pdf of $\text{Bin}(h, \beta)$, $\text{Bin}(i, \alpha)$ and q_k and giving the probability of going from states h (then) i to state j in a single step and can be set in (4) to yield the entropy rate of INAR(2) with P_{hi} as given by (5) and the stationary marginal pdf $\pi_i \equiv Pr(X_n = i)$.

Remark 5. Generally, suppose that X_1, X_2, \dots follow integer valued k th order autoregressive (INAR(k)) model

$$X_n = \alpha_1 \circ X_{n-1} + \alpha_2 \circ X_{n-2} + \dots + \alpha_k \circ X_{n-k} + \varepsilon_n, \quad n = k+1, k+2, \dots,$$

where nonnegative integer valued innovations ε_n 's following stationary pdf $q_i \equiv Pr(\varepsilon_n = i)$, $i = 0, 1, \dots$ are independent of all the binary counting series and X_{n-i} 's, $i = 1, 2, \dots, k$ (see, e.g., Jin-Guan and Yu(1991) for more details). Also, let $\pi_i \equiv Pr(X_n = i)$ be the stationary marginal pdf of the process. Then, by induction, the entropy rate of INAR(k) can be given by (3), where P_{ij} is the convolution pdf of the k binomial pdf's (with parameters $\alpha_1, \alpha_2, \dots, \alpha_k$) and the innovation pdf q_i .

Conclusion

In this paper, we consider the entropy rate of stationary Markov chains and integer valued autoregressive process as special case. The entropy rate of the process is the weighted sum of entropy values of transition probabilities which are the convolution of some binomial distributions and innovation distribution.

References

- Al-Osh, M.A. and Alzaid, A.A. (1987), First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis*, **8**, 261–275.
- Cover, T. M. and Thomas, J. A. Thomas (2006), *Elements of Information Theory*, John Wiley and Sons, New York.
- Dacunha-Castelle, D. and Duflo, M. (1994), *Probabilites et Statistiques, Problemes a temps mobile*, 2e edition, Masson, Paris.
- Jin-Guan, D. and Yu, L. (1991), The Integer Valued Autoregressive (INAR(p)), *Journal of Time Series Analysis*, **12(2)**, 129–142.
- McKenzie, E. (1985), Some simple models for discrete variate time series, *Water Resources Bulletin*, **21**, 645–650.
- Resnick, S. (2002), *Adventures in Stochastic Processes with Illustrations*, Springer, New York.
- Sadek, A. (2003), *Estimation des processus markoviens avec application en fiabilite*, These Univ. Techn. Compiegne.
- Steutel, F.W. and van Harn, K. (1979), Discrete analogues of self-decomposability and stability, *The Annals of Probability*, **7(5)**, 893–899.



On Distribution of the Sum of Dependent Random Variables

Najmeh Ahmadi¹ and Ali Dolati²

¹*Department of Statistics, College of Sciences, Shiraz University, 71454 Shiraz, Iran.*

stat.nahmadi@gmail.com

²*Department of Statistics, College of Mathematics, Yazd University, 89195-741, Yazd, Iran.*

adolati@yazd.ac.ir

Abstract: Mathematical probability theory interested to the computation of distribution of the sum of random variables. It is well known that the distribution of sum of random variables is the convolution of marginal distributions in the case of independence; no general characterization has been given to the general case of sum of dependent random variables. In this paper, we find expression for distribution of the linear combination of two dependent random variables in terms of their associated copula. We also study the effect of dependence structure on the distribution of them. Several examples illustrate our result.

Keywords Convolution, dependent random variables, Copula.

Mathematics Subject Classification (2010): 60E05.

1 Introduction

In actuarial applications, the insurance risks are shown by random variables which represent the amount of money that insurance company will have to pay out to the policyholder. These random amounts are depending on the occurrence of a claim. For example, in the case of a car insurance, the amount of a claim associated with the occurrence of an automobile accident is represented by a random variable. Random variables modelling insurance risks may generally be assumed to be non-negative. In many actuarial applications, the distributions of sum, product and ratio of two random variables need to be computed. For simplicity, the independence of random variables is assumed in

¹Speaker's Name : Najmeh Ahmadi Email:stat.nahmadi@gmail.com

many studies. In the absence of independence, we need methods and tools for modelling dependent variables. Copula functions are examples of such tools. When two random variables are dependent in most cases analytical solutions are not tractable and several techniques have been developed in literature to approximate or to investigate bounds for such distributions; see, e.g. [Arbenz et al. \(2011, 2012\)](#); [Frank et al. \(1987\)](#); [Gupta and Gupta \(2009\)](#); [Makarov \(1981\)](#); [Moynihan et al. \(1978\)](#); [Puccetti and Rüschendorf \(2012\)](#). In this paper, following the general approach addressed in [Cherubini et al. \(2011\)](#) we provide the expression for distribution of the linear combination of two dependent random variables, involving their copula and try to find solutions for some problems on distribution of sum of dependent random variables.

2 Preliminaries

Let (X, Y) be a pair of continuous random variables with the joint distribution function $H(x, y) = P(X \leq x, Y \leq y)$, univariate marginal distributions $F(x) = P(X \leq x)$, $G(y) = P(Y \leq y)$, at each $x, y \in \mathbb{R}$. Let C be the unique copula associated with (X, Y) through the relation

$$H(x, y) = C\{F(x), G(y)\}, \quad x, y \in \mathbb{R},$$

in view of *Sklar's Theorem* [Sklar \(1959\)](#). In fact, C is the cumulative distribution function of the pair $(U, V) = (F(X), G(Y))$ of uniform $(0,1)$ random variables. For a given C , let $\bar{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$ be the survival copula or reflected copula associated with C or equivalently, the cumulative distribution function of the pair $(1 - U, 1 - V)$. Note that $\bar{H}(x, y) = P(X > x, Y > y) = \bar{C}\{\bar{F}(x), \bar{G}(y)\}$, where $\bar{F}(x) = P(X > x)$ and $\bar{G}(y) = P(Y > y)$, are the survival functions of X and Y . Since copulas are Lipschitz continuous functions from $[0, 1]^2$ to $[0, 1]$ with Lipschitz constant equal to 1, then they are absolutely continuous in each argument [Nelsen \(2006\)](#), so that it can be recovered from any of its partial derivatives by integration. The partial derivatives of a copula C can be seen as conditional distribution functions $\dot{C}_1(u, v) = \partial C(u, v) / \partial u = P(V \leq v | U = u) \in [0, 1]$ and $\dot{C}_2(u, v) = \partial C(u, v) / \partial v = P(U \leq u | V = v) \in [0, 1]$. It follows that \dot{C}_1 will fail to be continuous on $(0, 1) \times [0, 1]$ if the distribution of V given $U = u$ has atoms. This phenomenon occurs for instance for the Fréchet lower and upper bound copulas, $W(u, v) = \max(u + v - 1, 0)$ and $M(u, v) = \min(u, v)$. We

denote by $\Pi(u, v) = uv$, the copula of two independent random variables. For cumulative distribution function F , let $F^{(-1)}(u) = \inf\{x|F(x) \geq u\}$, $u \in [0, 1]$, be the right continuous inverses of F .

3 Distribution of sums

In the following we study the distribution of the sum of two functions of dependent random variables. We assume that the random variables are continuous. We need the following lemma in the sequel.

Lemma 3.1. *Let X and Y be two continuous random variables with cumulative distribution functions F_X and F_Y . Let $\alpha(\cdot)$ and $\beta(\cdot)$ be two strictly increasing real-valued functions. Then for all t and $w \in [0, 1]$, the function*

$$\dot{C}_1(w, G \circ \beta^{-1}(t - \alpha \circ F^{-1}(w))),$$

is well defined almost surely (a.s.).

Theorem 3.2. *Let X and Y be two continuous random variables with the respective univariate marginal distribution functions F and G and the copula C . Let $\alpha(\cdot)$ and $\beta(\cdot)$ be two increasing real-valued functions and let $F \oplus_{\alpha, \beta}^C G$ be the distribution of $T = \alpha(X) + \beta(Y)$. Then*

$$F \oplus_{\alpha, \beta}^C G(t) = \int_0^1 \dot{C}_1(w, G \circ \beta^{-1}(t - \alpha \circ F^{-1}(w))) dw, \quad (3.1)$$

and the survival function of T , i.e., $P(T > t)$ is given by

$$\overline{F \oplus_{\alpha, \beta}^C G}(t) = \int_0^1 \dot{\bar{C}}_1(w, \bar{G} \circ \beta^{-1}(t - \alpha \circ \bar{F}^{-1}(w))) dw. \quad (3.2)$$

For the special $\alpha(x) = x$ and $\beta(y) = y$, (3.1) coincides with the C -convolution of F and G defined in Cherubini et al. (2011) by

$$F \oplus^C G(t) = \int_0^1 \dot{C}_1(w, G(t - F^{-1}(w))) dw. \quad (3.3)$$

Remark 3.3. *Note that if $\beta(\cdot)$ is decreasing then*

$$F \oplus_{\alpha, \beta}^C G(t) = \int_0^1 \dot{\bar{C}}_1(w, \bar{G} \circ \beta^{-1}(t - \alpha \circ \bar{F}^{-1}(w))) dw, \quad (3.4)$$

and

$$\overline{F \oplus_{\alpha,\beta}^C G}(t) = \int_0^1 \dot{C}_1(w, G \circ \beta^{-1}(t - \alpha \circ F^{-1}(w))) dw. \quad (3.5)$$

The operation $\oplus_{\alpha,\beta}^C$ defined by (3.1) is closed under convex combination. Let A and B two copulas and for $\lambda \in [0, 1]$, let $C = \lambda A + (1 - \lambda)B$. For all distribution functions F and G we have

$$F \oplus_{\alpha,\beta}^C G(t) = \lambda F \oplus_{\alpha,\beta}^A G(t) + (1 - \lambda)F \oplus_{\alpha,\beta}^B G(t).$$

Example 3.4. Let X and Y be two exponential random variables with the common cumulative distribution function $F(x) = 1 - e^{-\lambda x}$, $x > 0$. Then for three dependence structures Π , M and W , we have

$$F \oplus^{\Pi} F(t) = 1 - (1 + \lambda t)e^{-\lambda t}, \quad t > 0,$$

$$\begin{aligned} F \oplus^M F(t) &= \int_0^1 \mathbb{I}\{w : 2F^{-1}(w) \leq t\}(w)dw \\ &= \sup\{w \in (0, 1) : \frac{-\ln(1-w)}{\lambda} \leq \frac{t}{2}\} \\ &= 1 - e^{-\frac{\lambda}{2}t}, \quad t > 0, \end{aligned}$$

and

$$\begin{aligned} F \oplus^W F(t) &= \int_0^{F(t)} \mathbb{I}\{w : F^{-1}(w) + F^{-1}(1-w) \leq t\}dw \\ &= \int_0^{1-e^{-\lambda t}} \mathbb{I}\{w : w(1-w) \leq e^{-\lambda t}\}dw \\ &= \int_0^1 \mathbb{I}\{w : \frac{1 - \sqrt{1 - 4e^{-\lambda t}}}{2} \leq w \leq \min(1 - e^{-\lambda t}, \frac{1 + \sqrt{1 - 4e^{-\lambda t}}}{2})\}dw \\ &= \sqrt{1 - 4e^{-\lambda t}}, \quad t > \frac{2 \ln(2)}{\lambda}. \end{aligned}$$

Example 3.5. Let F and G be two uniform (0,1) distributions and let C be the FGM copula [Nelsen \(2006\)](#) given by

$$C(u, v) = uv(1 + \theta(1 - u)(1 - v)), \quad u, v \in [0, 1], \quad \theta \in [-1, 1]. \quad (3.6)$$

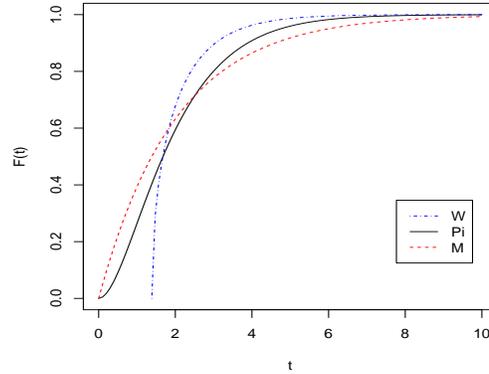


Figure 1: Plots of $X + Y$ for $X, Y \sim \text{Exp}(1)$ with $C = W, \Pi, M$.

Then

$$F \oplus^{FGM} G(t) = \begin{cases} 0, & t < 0 \\ \frac{t^2}{2} + \frac{\theta t^2(t^2 - 4t + 3)}{6}, & 0 \leq t < 1 \\ \frac{(-t^4 + 4t^3 - 3t^2 - 4t + 4)\theta}{6} - \frac{t^2}{2} + 2t - 1, & 1 \leq t < 2 \\ 1, & t \geq 2. \end{cases} \quad (3.7)$$

Note that (3.7) reduces to the distribution of the sum of two independent uniform $(0,1)$ random variables, i.e., the triangular distribution Johnson (1997, 1999), when $\theta = 0$. Thus, (??) could be considered as a generalized triangular distribution. The density function of (3.7) presented in Figure 2 for $\theta = -1, 1$.

If C_1 and C_2 are two copulas, we say that C_2 is more concordant than C_1 (written $C_1 \prec_c C_2$) if $C_1(u, v) \leq C_2(u, v)$ for all $(u, v) \in [0, 1]^2$. A totally ordered parametric family $\{C_\lambda\}$ of copulas is positively ordered if $C_{\lambda_1} \prec_c C_{\lambda_2}$ whenever $\lambda_1 \leq \lambda_2$. The copula C is said to be positive (resp, negative) quadrant dependent (PQD, resp, NQD) if $\Pi \prec_c (\succ_c) C$ Nelsen (2006). Note that if C is PQD then a pair (X, Y) with the copula C , are positively correlated. As a result, for all strictly

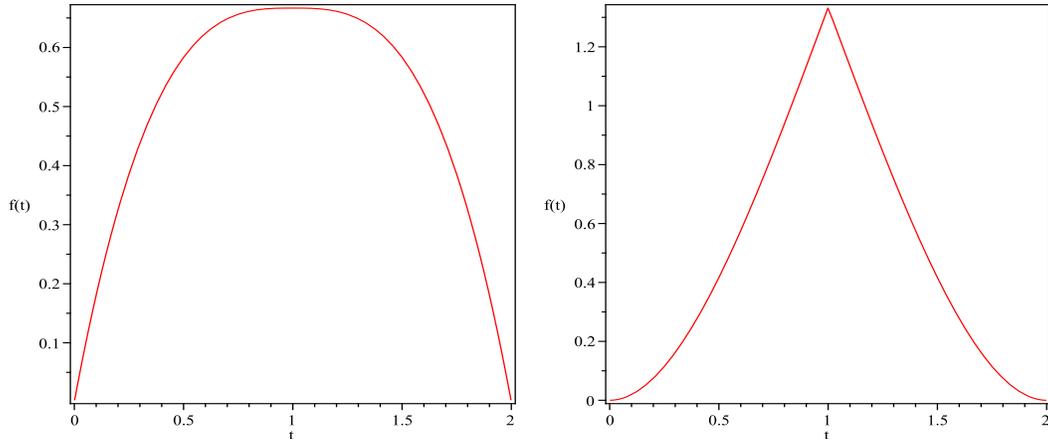


Figure 2: The density function of $X + Y$ in Example 3.5 with $\theta = 1$ (left panel) and $\theta = -1$ (right panel).

increasing transformations $\alpha(\cdot)$ and $\beta(\cdot)$, we have

$$\text{Var}(\alpha(X) + \beta(Y)) \geq \text{Var}(\alpha(X)) + \text{Var}(\beta(Y)).$$

The following result compares the variance of $\alpha(X) + \beta(Y)$ for members of a parametric family of copulas.

Theorem 3.6. *Let (X, Y) be a pair of continuous random variables with the copula C_λ , and let $\alpha(\cdot)$ and $\beta(\cdot)$ be two strictly increasing transformations on $\text{Ran } X$ and $\text{Ran } Y$, respectively. If C_λ is positively ordered, then $\text{Var}_\lambda(\alpha(X) + \beta(Y))$, is an increasing function of λ ; that is, for $\lambda_1 \leq \lambda_2$,*

$$\text{Var}_{\lambda_1}(\alpha(X) + \beta(Y)) \leq \text{Var}_{\lambda_2}(\alpha(X) + \beta(Y)).$$

Since for every copula $W \prec_c C \prec_c M$, for every copula C , we have the following result.

Corollary 3.7. *For every copula C , if $\alpha(\cdot)$ and $\beta(\cdot)$ are two increasing functions on $\text{Ran } X$ and $\text{Ran } Y$, respectively, then*

$$\text{Var}_W(\alpha(X) + \beta(Y)) \leq \text{Var}_C(\alpha(X) + \beta(Y)) \leq \text{Var}_M(\alpha(X) + \beta(Y)),$$

where

$$\text{Var}_W(\alpha(X) + \beta(Y)) = \text{Var}(\alpha \circ F^{-1}(U)) + \text{Var}(\beta \circ G^{-1}(U)) + 2\text{Cov}(\alpha \circ F^{-1}(U), \beta \circ G^{-1}(1 - U)),$$

and

$$\text{Var}_M(\alpha(X) + \beta(Y)) = \text{Var}(\alpha \circ F^{-1}(U)) + \text{Var}(\beta \circ G^{-1}(U)) + 2\text{Cov}(\alpha \circ F^{-1}(U), \beta \circ G^{-1}(U)),$$

where U is a uniform $(0,1)$ random variable.

Example 3.8. Let X and Y be two exponential random variables with the parameter λ and the associated copula C . Then

$$\text{Var}_W(X + Y) = \frac{(4 - \pi^2/3)}{\lambda^2}, \quad \text{Var}_M(X + Y) = \frac{4}{\lambda^2}, \quad \text{Var}_\Pi(X + Y) = \frac{2}{\lambda^2}.$$

If the copula of C is PQD then

$$\frac{2}{\lambda^2} \leq \text{Var}_C(X + Y) \leq \frac{4}{\lambda^2},$$

and if C is NQD then

$$\frac{(4 - \pi^2/3)}{\lambda^2} \leq \text{Var}_C(X + Y) \leq \frac{2}{\lambda^2}.$$

Example 3.9. Consider the distribution of the sum of two dependent uniform $(0,1)$ random variables with copula FGM, given in (3.7). The variance of this distribution is equal to $(\theta + 3)/18$ and limited to the interval $[1/9, 2/9]$. Note that a well-known limitation to the FGM family of copulas is that it does not allow the modeling of large dependencies since, for example, Spearman's rho [Nelsen \(2006\)](#) of this family is limited to $[-1/3, 1/3]$.

References

- Arbenz, P., Embrechts, P. and Puccetti, G. (2011). AEP algorithm for the fast computation of the distribution of the sum of dependent random variables. *Bernoulli*, **17(2)**, 562–591.
- Arbenz, P., Embrechts, P. and Puccetti, G. (2012). The GAEP algorithm for the fast computation of the distribution of a function of dependent random variables. *Stochastics*, , 569–597.
- Cherubini, U., Mulina, S. and Romagnoli, S. (2011). On the distribution of the (un)bounded sum of random variables. *Insurance: Mathematics and Economics*, **48**, 56–63.

- Cossette, H., Cote, M-P., Mailhot, M. and Marceau, E. (2014). A note on the computation of sharp numerical bounds for the distribution of the sum, product or ratio of dependent risks. *J. Multivariate Analysis*, **130**, 1–20.
- Darsow, W.F., Nguyen, B. and Olsen, E.T. (1992). Copulas and Markov processes. *Illinois J. Math.* **36**, 600–642.
- Frank, M.J., Nelsen, R. and Schweizer, B. (1987). Best-possible distributions of a sum—a problem of Kolmogorov. *Probability theory and related Fields*, **74**, 199–211.
- Gupta, P.L. and Gupta, R. C. (2009). Distribution of a linear function of correlated ordered variables. *Journal of Statistical Planning and Inference*, **139**, 2490–2497.
- Hoeffding, W. (1940). Scale-invariant correlation theory. In: Fisher, N.I., Sen, P.K. (Eds.), *The Collected Works of Wassily Hoeffding*. Springer-Verlag, New York, 1994.
- Johnson, D. (1997) The triangular distribution as a proxy for the beta distribution in risk analysis. *Statistician*, **46**, 387–398.
- Johnson, N. L. and Kotz, S. (1999) Non-smooth sailing or triangular distributions revisited after some 50 years. *Statistician*, **48**, 179–187.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London
- Makarov, G.D. (1981). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of probability and its Applications*, **26**, 803–806.
- Moynihan, R., Schweizer, B. and Sklar, A. (1978). Inequalities among binary operations on probability distribution functions. *General Inequalities*, **1**, 133–149.
- Nelsen, R.B. (2006). *An Introduction to Copulas. Second Edition*. Springer, New York.
- Puccetti, G. and L. Rüschendorf (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *J. Comp. Appl. Math.*, **236**, 1833–1840.

Rohatgi, V.K. (1976). *An Introduction to Probability Theory Mathematical Statistics*. Wiley, New York.

Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.

Springer, M.D. (1979). *The Algebra of Random Variables*. Wiley, New York.



A Bayesian method for bandwidth choice with length-biased data

Masoud Ajami bakhtyarvand ¹

Vali-e-Asr University of Rafsanjan

Abstract: In sampling, arisen data with probability proportional to its length is called Length-biased . Nonparametric density estimation in length-biased sampling is more difficult than other states . A local bandwidth selection method in the smoothed kernel estimation framework was performed in length-biased data by . In this paper , we introduce a new estimator , supposing the bandwidth as a random variable , following inverted gamma prior distribution and it is estimated in Bayesian approach with the lognormal kernel that is proved the pointwise strong consistency of this estimator . Then , the results obtained from the simulation study , carried out to assess a new proposed estimator is compared with the estimator shown by . The noticeable result in this study is that the proposed estimation perform better than that estimator .

Keywords Bayesian Approach , Kernel Density Estimator , Lognormal Kernel , Local Bandwidth Parameter , Monte carlo method .

Mathematics Subject Classification (2010) : 62Fxx.

1 Introduction

Length-biased data arise in many practical situations , including econometrics , survival analysis , renewal processes , biomedicine and physics . For instance , if represents the length of an item and the probability of this item selected in the sample is proportional to its length , then the distribution of the observed length is length-biased . Interesting applications of length-biased data can be found in Cox (1969) . Given a distribution function (d.f.) F , we say a random variable (r.v.) Y has the length-biased distribution of F if the d.f. of Y is given by

$$G(t) = \int_0^t \frac{x dF(x)}{\mu}, \quad t \geq 0, \quad (1.1)$$

where $\mu = \int_0^\infty x dF(x)$, and μ is assumed to be finite . In the case that F has a density f with respect to the Lebesgue measure , (1.2) can be written as

$$G(t) = \int_0^t \frac{x f(x) dx}{\mu}, \quad t \geq 0. \quad (1.2)$$

An elementary calculation shows that F is determined uniquely by G , namely

$$F(t) = \mu \int_0^t y^{-1} dG(y), \quad t \geq 0.$$

Let Y_1, \dots, Y_n be a sample from G . The empirical estimator of can be written in the form of

$$F_n(t) = \mu_n \int_0^t y^{-1} dG_n(y), \quad (1.3)$$

where

$$\mu_n^{-1} = \int_0^\infty y^{-1} dG_n(y).$$

G_n is an empirical estimator of G given by

$$G_n(t) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq t),$$

where $I(A)$ denotes the indicator of the event A . [Bhattacharyya et al . \(1988\)](#) based on kernel approach proposed the following estimator for length-biased data Y_1, \dots, Y_n

$$f_n^*(t) = \frac{\mu_n}{t} \sum_{i=1}^n \frac{1}{nh_n} K\left(\frac{t - Y_i}{h_n}\right). \quad (1.4)$$

Here $K(\cdot)$ is a symmetric kernel and $\{h_n, n \geq 1\}$ is a sequence of positive bandwidths satisfying $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ as $n \rightarrow \infty$, which controls the degree of smoothing . They proved consistency and asymptotic normality for f_n^* . Later , [Jones \(1991\)](#) proposed the following estimator

$$f_n(t) = \frac{\mu_n}{nh_n} \sum_{i=1}^n \frac{1}{Y_i} K\left(\frac{t - Y_i}{h_n}\right). \quad (1.5)$$

As mentioned by Jones (1991), this estimator has various advantages with respect to (1.4). It is a probability density function. Overall estimator for favorite kernel density function (symmetric or asymmetric) has been suggested below

$$\hat{f}_h(x) = n^{-1} \hat{\mu}_n \sum_{j=1}^n Y_j^{-1} K(x; Y_j, h). \quad (1.6)$$

Two important factors of kernel density and bandwidth parameter have important roll for kernel density estimation. It is of note that selection of the bandwidth parameter is more important than selection of kernel function. There are several references in this case that can mention to Silverman (1978).

In the estimation discussion of the density function, the bandwidth parameter is unknown that should be estimation. There are several bandwidth selection methods available for reader such as least square cross validation, nearest neighbor and biased cross validation. In all of these methods bandwidth parameter h is considered as a global parameter. With censoring data Kulasekera and Padgett (2006) estimator of $h(x)$ in the mean posterior, under the mean squared loss function by appropriate prior selection for local bandwidth. Also Ajami et al. (2011) investigated Bayes bandwidth in length-biased data.

In this paper, the next section deals with the study of bandwidth selection in Bayesian method by using asymmetric lognormal in (1.6) proved its pointwise strong consistence. In section 3, with the simulation study we compare the proposed lognormal kernel density estimator $\hat{f}_{h_L}(t)$ with another kernel density estimator $\hat{f}_{h_I}(t)$ using an inverse Gaussian kernel that was shown by Ajami et al (2011). The both estimators use Bayesian local bandwidths associated with their respective kernels.

2 Bayesian method in bandwidth estimator

In this section, bandwidth parameter (h) is assumed as a random variable that the data are length-biased and the density function f is estimated with the kernel density estimator (1.6). Suppose that $\pi(h)$ is a prior distribution with considering of random sample Y_1, \dots, Y_n from length-biased distribution G . The posterior density of h will be followed as

$$\hat{\pi}(h|Y_1, \dots, Y_n, x) = \frac{\hat{f}_h(x)\pi(h)}{\int \hat{f}_h(x)\pi(h)dh}, \quad (2.1)$$

where $\hat{f}_h(x)$ is (1.6) .

In below theorem , the bandwidth of (1.6) is given locality based on lognormal kernel and inverted gamma prior distribution .

Theorem 2.1. *Suppose that kernel density is Lognormal with parameters h and $\ln Y_j$ which is defined as*

$$K(x, h, \ln Y_j) = \frac{1}{\sqrt{2\pi}} \frac{1}{xh} e^{-\frac{1}{2} \left(\frac{\ln x - \ln Y_j}{h} \right)^2}, \quad (2.2)$$

and the bandwidth h follows an inverted gamma prior distribution with parameters α and β

$$\pi(h) = \frac{2}{\beta^\alpha \Gamma(\alpha) h^{2\alpha+1}} e^{-\left(\frac{1}{\beta h^2}\right)}, \quad \alpha > 0, \beta > 0. \quad (2.3)$$

Then under the squared error loss , the Bayesian local bandwidth h for estimating at x is given by

$$h_L = h(x) = E[(h|Y_1, \dots, Y_n, x)] = \frac{\sum_{j=1}^n Y_j^{-1} (\beta_j^*)^{(\alpha^*)}}{(\alpha^*) \sum_{j=1}^n Y_j^{-1} (\beta_j^*)^{(\alpha^*)}}, \quad (2.4)$$

where β_j^* and α^* will be defined in proof . Also when $n \rightarrow \infty$ we have

$$h_L = h(x) = E[(h|Y_1, \dots, Y_n, x)] \rightarrow 0 \quad a.s. \quad (2.5)$$

Proof. By replacing (2.2) in (1.6) and according to (2.1) it can be writing as

$$\hat{\pi}(h|Y_1, \dots, Y_n, x) = \frac{\sum_{j=1}^n Y_j^{-1} \frac{2}{(h^2)^{\alpha+1}} e^{-\frac{1}{h^2 \beta_j^*}}}{\Gamma(\alpha^*) \sum_{j=1}^n Y_j^{-1} (\beta_j^*)^{\alpha^*}},$$

where

$$\beta_j^* = \left[\frac{1}{\beta} + \frac{(\ln x - \ln Y_j)^2}{2} \right]^{-1}$$

and

$$\alpha^* = \alpha + \frac{1}{2}, \alpha > \frac{1}{2}.$$

Then , the Bayes estimator of h at x is calculated as the mean of this posterior

$$h_L = \frac{\sum_{j=1}^n Y_j^{-1} (\beta_j^*)^{(\alpha)}}{\Gamma(\alpha^*) \sum_{j=1}^n Y_j^{-1} (\beta_j^*)^{(\alpha^*)}}.$$

Because the prof of (2.5) likes theorem 4.1 in [Kuruwita et al. \(2010\)](#) we don't proof it .

□

Theorem 2.2. *Let $Y_1 \dots Y_n$ is a sample of length-biased distribution G and $G(\tau_F) > 0$, where $\tau_F = \inf\{\tau : 1 - F(t) > 0\}$. Suppose f is a bounded density then for each $0 < x < \tau$ and $h_L = h_n(x) \rightarrow 0$ when $n \rightarrow \infty$, we have*

$$|\hat{f}_{h_L}(x) - f(x)| \xrightarrow{a.s.} 0.$$

Proof. Suppose G_n ia a empirical distribution function of Y_i 's , then F

$$F_n(t) = \hat{\mu}_n \int_0^t y^{-1} dG_n(y).$$

So

$$\hat{f}_{h_L}(x) = \int_0^\infty K(x; u; h_L) dF_n(u). \quad (2.6)$$

Let

$$f_{h_L}(x) = \int_0^\infty K(x; u; h_L) dF(u).$$

Then

$$\begin{aligned} |\hat{f}_{h_L}(x) - f(x)| &\leq |\hat{f}_{h_L}(x) - f_{h_L}(x)| + |f_{h_L}(x) - f(x)| \\ &=: J_1 + J_2. \end{aligned} \quad (2.7)$$

Now we have

$$J_1 \leq \sup_{0 < t \leq T} |F_n(t) - F(t)| \left| \int_0^\infty dK_u(x; u; h_L) \right|. \quad (2.8)$$

According to theorem 4.2 . of [Horváth , L . \(1985\)](#) can be showed that F_n is a uniformly consistent for $F(t)$. So if $n \rightarrow \infty$ than

$$\sup_{0 < t \leq T} |F_n(t) - F(t)| \rightarrow 0, \quad a.s.$$

On the overhand whit some calculation we have

$$\left| \int_0^\infty dK_u(x; u, hL) \right| \leq \frac{1}{xh^*}.$$

Then if $n \rightarrow \infty$ $J_1 \rightarrow 0$. Finalle like theorem 1 [Kuruwita et al. \(2010\)](#) can be showed $J_2 \rightarrow 0$. So the proof is complete . \square

3 Simulation

In this section , the effect of the kernel is discussed on the proposed estimator . In particular , the proposed lognormal kernel density estimator is compared with that of shown by [Ajami et al . \(2011\)](#) that both estimators using Bayesian local bandwidths are associated with their respective kernels . Also in both estimators , bandwidth h follows inverted gamma prior distribution with parameters α and β . The performance of the two kernels , namely the lognormal and the inverse Gaussian , is compared with 2 different sample sizes $n = 25$ and $n = 50$. In all simulation for Bayes bandwidth computation it is used an inverted gamma prior with $\alpha = 2$ and $\beta = 2$, and is considered that the data are emanated from an unbiased model with underlying Gamma density function

$$f(t) = \frac{t^{\gamma-1} e^{-\frac{t}{\theta}}}{\theta^\gamma \Gamma(\gamma)}, \gamma > 0, \theta > 0, t \geq .$$

Thus the length-biased density is

$$g(t) = \frac{t^\gamma e^{-\frac{t}{\theta}}}{(\theta)^{\gamma+1} \Gamma(\gamma+1)}, \gamma > 0, \theta > 0, t \geq .$$

The pointwise estimated mean squared error (*EMSE*) of any density estimator $\hat{\phi}_n(t)$ is defined at a point of estimation t , by

$$EMSE(\phi_n(t)) = \frac{1}{N} \sum_{i=1}^N (\phi_n(t) - \phi(t))^2,$$

where N is the number of simulations which was chosen to be 1000 and $\phi(t)$ ia any kernel density . All simulations are carried out using Mont carlo method with R . Then , the ratio is examined

$$r(t) = \frac{EMSE(\hat{f}_{h_L}(t))}{EMSE(\hat{f}_{h_I}(t))},$$

over a grid of t values in the domain of the underlying density . These ratios are plotted against t to assess the pointwise performance of the two density estimators $\hat{f}_{h_L}(t)$ and $\hat{f}_{h_I}(t)$.

Density estimates of a unbiased density $Gamma(\gamma, 1)$ are computed with use of the two kernel density estimators (LN and IG) and the ratio $r(t)$ is plotted against the domain values t of the underlying density . As shown in Figure 1 , the lognormal kernel density estimator is clearly weaker than the inverse Gaussian kernel density estimator in the neighborhood of the origin . However , the decreasing ratio values indicate that the pointwise estimator for the lognormal kernel density estimator are not as close as the ones is gotten from the inverse Gaussian kernel density estimator toward the end of the support . The increasing sample size shows no significant effect on the performance of the kernel density estimators .

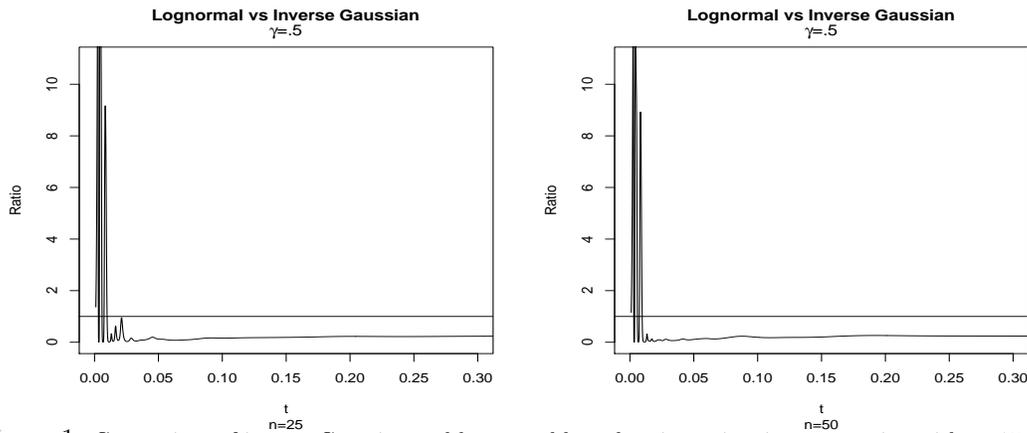


Figure 1: Comparison of inverse Gaussian and lognormal kernels using pointwise error ratio with $n=25$, $n=50$.

Next , a similar pattern is observed in the performances of the two density estimators as with the last positions of $gamma = 1$. The lognormal kernel density estimator is far superior than the inverse Gaussian kernel density estimator in the neighborhood of the origin . Further , the proportion of domain values in which the lognormal kernel density estimator is superior has increased with the

increment of the sample size as indicated in Figure 2 .

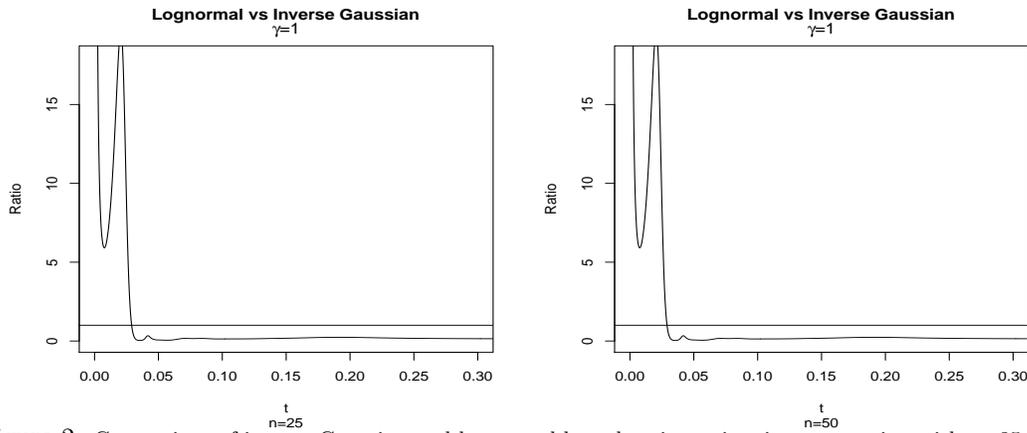


Figure 2: Comparison of inverse Gaussian and lognormal kernels using pointwise error ratio with $n=25$, $n=50$.

According to Figure 3 , we observe that the previous two settings are alike with the all cases . The most noticeable feature in this case is that the lognormal kernel density estimator performe better than the inverse Gaussian kernel density estimator over a large proportion of the support with the increment of the sample size .

References

- Ajami , M . , Fakoor , V . , and Jomhoori , S . (2011) . Bayesian and cross validation estimation bandwidth of kernel density function estimator for length-biased data . *Journal of Statistical Sciences(in Persian)***5**, 41–60 .
- Bhattacharyya , B . B . , Franklin , L . A.& Richardson , G . D . (1988) . A comparison of nonparametric unweighted and length-biased density estimation of fibres . *Comm . Statist A* **17**, 3629–3644 .
- Cox , D . R . (1969) . *Some sampling problems in technology . New developments in survey sampling*, (eds : N.L . Johnson and H . Smith) . New York : Wiley .
- Horváth , L . (1985) . Estimation from a length-biased distribution. *Statistics and Decisions* **3**, 91–113.

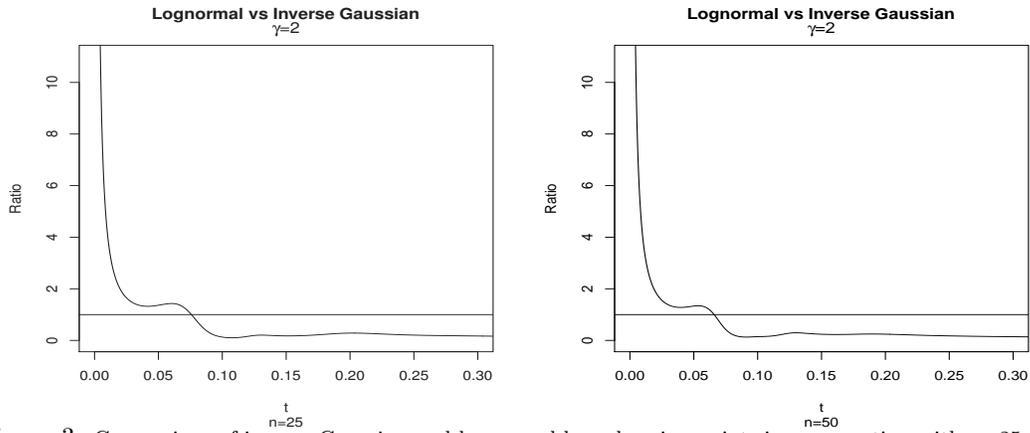


Figure 3: Comparison of inverse Gaussian and lognormal kernels using pointwise error ratio with $n=25$, $n=50$.

Jones , M . C . (1991) . Kernel density estimation for length biased data. *Biometrika* **78**, 511–519 .

Kulasekera , K . B. , and Padgett , W.J . (2006) . Bayes bandwidth selection in kernel density estimation with censored data . *Nonparametric Statistics* , **18**, 129–143 .

Kuruwita , C . N and Kulasekera , K . B and Padgett , W . J.(2010) . Density estimation using asymmetric kernels and Bayes bandwidths with censored data . *Journal of Statistical Planning and Inference*,**7**, 1765–1774 .

Silverman , B . M . (1978) . Weak and strong uniform consistency of the kernel estimate of a density and its derivatives . *Ann . Statist* . **6** , 177–184 . **27**, 832–837 .



Survey to Gene Network Interaction in Microarray Data Analysis

H. Alavi majd¹, A. Talebi ¹, A. Dolati ³, K. Gilani⁴

¹ Professor of Shahid Beheshti University of Medical Sciences, Department of Biostatistics, School of Paramedical Sciences, Tehran, Iran

² PhD Student of Biostatistics Department, School of Paramedical Sciences, Students research committee, Shahid Beheshti University of Medical Sciences, Tehran

³ Assistant Professor of Yazd University, Department of Statistics, Yazd

⁴ Reproductive Biotechnology Research Center, Avicenna Research Institute, ACECR, Tehran

Abstract: Many biological research fields such as cellular and molecular function design need to gene regulatory networks to provide obvious insight and understanding of the biological process, molecular function and cellular component in living cells. These interactions among the genes and their products play an important role in molecular processes. Because of its importance, several computational techniques have been suggested to infer gene regulatory networks from gene expression data. In this review study, three inference methods are discussed: Bayesian network, Boolean network, Correlation coefficient methods. These approaches are discussed in terms of introduction, methodology and recent applications of these approaches in gene regulatory network construction. These approaches are also compared in the discussion section. Moreover, the strengths and weaknesses of these computational techniques are described.

Keywords Gene regulatory network, Bayesian network, Boolean network, Pajek software.

Mathematics Subject Classification (2010): 99X99 99X99 99X99.

¹Speaker's Name : Atefe Talebi a.talebi@sbmu.ac.ir

1 Introduction

1.1 Bayesian network methodology

Bayesian network is a directed and acyclic graph where $G(V, E)$ with a set of local probability distributions P . V displays the nodes, $\{V_1; V_2; \dots; V_N\}$, which explain the gene variables (1). E indicates the directed edge that corresponds to probabilistic dependence interactions among genes. A DAG is a network graph with directed edges and without cycles. In a DAG, the edge pointing from one node to another displays the regulatory relationship between parent node and child node. For instance, we assume V_1 is the parent of V_2 and V_2 is the parent of V_3 , the assumptions are that V_1 is the ancestor of V_3 and V_3 is the descendant of V_1 . The interaction between two nodes linked by an undirected edge is described by the neighbor term or adjacent term (2). The joint distribution of a set of variables is represented as:

$$P(v_1, v_2, \dots, v_n) = \prod_{(i=1)}^n P(v_i | \text{parents}(v_i))$$

where v_i is a given gene node. n is the total number of genes involved. The $\text{parents}(v_i)$ is the parent gene that regulates gene v_i . $P(v_i | \text{parents}(v_i))$ for node i is represented as Conditional Probability Distribution or local distribution. Conditionally Probability Distribution must be specified for the nodes that have parents. Peer et al in 2001 studied the Bayesian network framework to deal with perturbations in gene regulatory network inference (3). Also, a discretisation process is applied to the data prior to gene regulatory network inference. First, the data were considered as a discretisation process. These steps included inferring gene regulatory network using Bayesian Network, and finding the optimal gene network according to the scoring function. The intervention is not surveyed by using genetic mutation experiments but by substituting external molecular function to the original one based on the expression of V . The results proposed new characteristics and displayed the differences between the proposed technique, the clustering and gene set techniques.

Werhli and Husmeier in 2007 combined expression data with the prior biological knowledge from multiple resources to construct gene regulatory network (4). They applied Bayesian network with the information from expression data and prior biological knowledge to improve the prediction accuracy.

In addition to, they used Markov Chain Monte Carlo (MCMC) sampling scheme to sample the hyperparameters from posterior distribution. They concluded that the values of the hyperparameters were close to optimal for reducing the construction error.

Tanand Mohamad in 2012 studied a new Bayesian network algorithm to make gene regulatory networks with the hill-climbing and Efron's bootstrap methods (5). At the first step, the Least Local Squares (LLS) imputation algorithm was used to impute the missing values that existed in the microarray data set. After that, Bayesian network was used for network model construction and learned by using hill-climbing algorithm. The bootstrap method was applied to extract a set of edges with a high confidence level. They deduced that the gene regulatory network achieved a high true positive rate, and more novel relationships among the genes were shown by their method.

Yang et al in 2011 proposed reducing the computational time of Bayesian network using the Sparse Graph Search (SGS) approach (6). The SGS algorithm applies iterated statistical independence test and searching technique to find the optimal network structure. Moreover, the optimal gene regulatory network is produced using a hybrid method. The proposed hybrid approach contain the search-and-score and constraint-based methods. For the search-and-score approach, a scoring metric evaluates the network quality based on the dataset and searches the best network structure overall candidate networks using the optimization method. However, the constraint-based approach surveys the existence of an edge by applying conditional independency tests instead of using statistical or information theory measurements. Their results displayed that the accuracy and computational efficiency was improved by their approach.

1.2 Boolean network

Boolean networks are the simplest models to construct gene regulatory networks (7). The method is easy to simulate and popular to capture the global dynamical behavior of genetic regulatory networks and interactions of genes (8). Correlation-based and inferring-based are two types of inferring algorithms that are used in the Boolean networks (9). For the correlation approach, gene relationship information is obtained by different methods, which are then used in modeling the topology of the

relationships between genes. The inferring-based method is a machine learning method and the most commonly used algorithm in genetics. A Boolean is defined as variable that can only consider two values; the values are usually displayed by 1 and 0 (true and false). The logic operators are and, or and not. For example, $x_1; x_2$ and $x_3 = 1$ means that if and only if their variables are 1. The interaction types are separated into two classes for building a model: inactive and active (10). A Boolean function is a function in which Boolean variables are connected by logic operators by means that, the situation of gene expression level is either on or off in a Boolean network.

A Boolean network is a directed graph $G(V,E)$ and represented by variable $v \in V$, whereas the node or variable is related to the function in Boolean. The state of the network is shown that by values of nodes where S is denoted as the state, $S(t) = (x_1(t), x_2(t), \dots, x_n(t))$. All nodes are updated according to the function $x_i(t+1) = b_i(x_{i1}(t), x_{i2}(t), \dots, x_{in}(t))$ (11).

One of the most attractive aspects of Boolean network is their simplicity, fast and efficient, although they are not suitable in big data. Furthermore, Boolean network apply time-discrete and synchronous updates. Also, the limitations of Boolean network are that they are not able to capture many important details of system behavior. The deterministic nature of Boolean network is a problem because noisy data in the computation may cause inconsistency (12).

Akutsu et al in 1999 studied the Boolean network of a genetic network from the situation of a transition table with a time series of gene expression (13). The results displayed that small numbers of the transition state pair are sufficient in inferring the original Boolean network. A simpler algorithm has been extended which represented that if the in degree of each node is bounded by a constant, $O(\log n)$, the $O(\log n)$ state transition pairs are sufficient to determine the original Boolean network. The constant factor of $O(\log n)$ is exposed. It has been displayed that Boolean network with a size of 100,000 can be identified by 100 INPUT/OUTPUT pairs only if the maximum in degree is 2. The approach showed a simple and extensible method for realistic network models.

Saadatpour and Albert in 2013 proposed that Boolean logic can be applied to describe the interactions between mRNAs and proteins to predict gene patterns, especially for a gene that has not been studied yet (14). mRNA or protein is shown by a node within the network, while the edge displays the interaction between them. The state is either 1 or 0 based on the presence or absence of one substance.

The state of a node varies over time. Time intervals that are greater than or equal to the period of central dogma, DNA makes RNA and RNA makes protein, are used. The Boolean function determines the situation of node i . The results showed that as long as their net effect is preserved, the kinetic details of the interactions do not matter. Through adding a minimum of additional kinetic hypotheses, a Boolean network permits the integration of qualitative observations on gene interactions into a coherent picture.

1.3 Correlation coefficient methods

Butte and Kohane in 2000 proposed a method that calculated pair-wise mutual information technique in *Saccharomyces cerevisiae* data set (15). They selected 2467 RNA expression data to construct 22 clusters, and they explained biological relationship of each clusters or Relevance Networks. Clusters were chosen by a fold-change greater than threshold criteria. They calculated the entropy of gene expression patterns and also the mutual information between RNA expression patterns for each gene pairs. High entropy showed that its expression levels are randomly distributed, and high mutual information between two genes showed that one gene was non-randomly connected to each other gene. In the article, there were 143 genes, 102 various components of the large and small ribosomal subunits, and 8 translation initiation factors in the largest cluster.

Cheng et al in 2013 suggested extracting non-linear relationships that is called Kernel correlation coefficient and comparing with Pearson correlation coefficient and Mutual information method in *Saccharomyces cerevisiae* data (16). They construct a correlation matrix across the entire set of microarray data set and construct gene networks by using the graph theory. Although our approach is described for and applied to data from the yeast microarray, our method is equally applicable to data from the Affymetrix or other array platforms, or for other types of high-throughput quantitative data. The reliability of their proposed method was checked with DREAM4. The ROC and PR were used to assess three methods separately. AUROC of the Gaussian kernel correlation coefficient was the largest in all. The interaction genes network is displayed graphically by using the Pajek software (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>).

References

- Akutsu, T., Miyano, S. and Kuhara, S. (1999). Identification of genetic networks from a small number of gene expression pattern under the Boolean model. *Pacific Symp Biocomput* , **4**, 17-28.
- Ando, S., Sakamoto, E. and Iba, H. (2002). Evolutionary modeling and inference of gene network. *Inf.Sci* , **145**, 237-259.
- Butte, A.J., Kuhara, I.S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput* , **5**, 418-423.
- Cheng, L., Khorasani, K., Ding, Y and Guo, X. (2013). Gene interaction networks based on kernel correlation metrics. *Int J Comput Biol Drug Des* , **6**, 72-92.
- Chen, C.C., Zhong, S. (2008). Inferring gene regulatory networks by thermo dynamic modelling. *BMC Genomics*, **9**, S19-S25.
- Davidich, M.I., Bornholdt, S. (2008). Boolean network model predicts cell cycle sequence of fission yeast. *PLoSOne* , **3**, e1672-1675.
- Kaderali, L. and Radde, NL. (2008). Inferring gene regulatory networks from expression data. *Stud Comput Intell* , **94**, 33-74.
- Lahdesmaki, H., Shmulevich, I. and Yli-Harja, O. (2003). On learning gene regulatory networks under the Boolean network. *Model Mach.Learning* , **52**, 147167.
- Lee, W.P., Tzou, W.S. (2009). Computational methods for discovering gene networks from expression data. *Brie.Bioinf* , **10**, 408-423.
- Mihajlovic, V. and Petkovic, M. (2001). Dynamic Bayesian Networks: A State of the Art, CTIT Technical Report Series. *TR-CTIT* , **34**, 1-37.
- Pe'er, D., Regev, A. and Elidan, G. Friedman, N (2002). Inferring sub networks from perturbed expression profiles. *Bioinformatics* , **17**, S215-S224.

- Saadatpour, R., Albert, A. (2013). Boolean modeling of genetic regulatory networks:a methodology tutorial. *Methods* , **62**, 3-12.
- Tan, A.K., Mohamad, M.S. (2012). Using Bayesian networks to construct gene regulatory networks from microarray data. *J.Teknol* , **58**, 1-6.
- Yang, B., Zhang, J. and Li, A. (2011). A Bayesian Network based algorithm for gene regulatory network reconstruction. *signal processing Commun Comput IEEE Int Conf* , frn[o]--4.
- Velculesc, V.E., Zhang, M.S., Vogelstein, B and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* , **270**, 484487.
- Werhli, AV., Husmeier, D. (2007). Reconstructing gene regulatory networks with bayesian networks by combining expression data with multiple sources of prior knowledge. *Stat. Appl. Genet. Mol. Biol* , **6**, 15.



An asymmetric Markov switching GARCH model: estimation and forecasting

N. Alemohammad¹, S.Rezakhah² and S. Hosseinalizadeh^{3 2}

¹Department of Mathematics and Computer Science, Shahed University, Tehran,

²Faculty of Mathematics and Computer Science, Amirkabir University of Technology,
Tehran, Iran.

³Department of Computer Engineering and Information Technology, Qazvin Branch,
Islamic Azad University, Qazvin, Iran

Abstract: In this paper an extension of Markov switching GARCH model is proposed to model leverage effect, conditional heteroscedasticity and nonlinearity attributes of financial time series. Greedy Gibbs Bayesian learning method is used to estimate the parameters of the model. Due to the complexity of the model a dynamic programming algorithm for forecasting is proposed. Finally we illustrate the model on S&P500 daily returns.

Keywords: GARCH models, Markov switching, Forecasting, Bayesian inference, Griddy Gibbs sampling.

Mathematics Subject Classification (2010): 60J10, 62M10, 62F15.

1 Introduction

In the past four decades, dynamic financial time series analysis based on nonlinear models has become a topic of interest for some researchers. For financial time series, the ARCH model and GARCH model are surely the most popular classes of volatility models. Merging (G)ARCH model with a hidden Markov chain, where each state of the chain allows a different (G)ARCH behavior, was introduced by Hamilton. Gray (1996), Haas et al. (2012) proposed different variants of Markov-Switching GARCH

²N. Alemohammad : n.alemohammad@shahed.ac.ir

(MS-GARCH) model and its applications. Also some other researchers such as [Bauwens et al. \(2010\)](#), [Billio et al. \(2014\)](#) have done extensive studies on the properties of this model.

One restitution of the GARCH model is the symmetry in the response of volatility to past shocks (because the volatility is a function of the lagged squared observations). In other words, the dynamics of conditional variance in (G)ARCH models change only with the size of squared observations. Financial markets become more volatile in response to negative shocks than to positive shocks. The asymmetric GARCH model (AGARCH) introduced by [Engle \(1990\)](#), Threshold GARCH (TGARCH) model of [Zakoian \(1994\)](#) and Skew normal mixture and Markov-switching GARCH [Haas \(2010\)](#) capture the asymmetric effects of negative and positive shocks.

In this paper, we propose a new structure for the Markov switching GARCH model where the volatility of each regime has different reactions to positive and negative shocks. This model is talented to switch between regimes with distinct levels of volatilities and also encompasses the leverage effect in each regime. By considering different smooth transition functions for each state of Markov chain (to capture asymmetry property in each state), our model is able to have more accurate forecast than the MS-GARCH. Since using all past observations for forecasting could increase the complexity of the model, we reduce the volume of calculations by proposing a dynamic programming algorithm. For the estimation of the parameters, the Bayesian inference is used.

The organization of this paper is as follows: in the next section the structure of our model is presented. Section 3 is devoted to the estimation of the parameters of the model. Section 4 is dedicated to the analysis of the efficiency of the proposed model via an empirical application and the comparison of forecast errors with the MS-GARCH model. Section 5 concludes.

2 Markov switching smooth transition GARCH model

The Markov switching smooth transition GARCH model, MS-STGARCH, for time series $\{y_t\}$ is defined as

$$y_t = \varepsilon_t \sqrt{H_{t,Z_t}}, \quad (2.1)$$

where $\{\varepsilon_t\}$ are iid standard normal variables, $\{Z_t\}$ is an irreducible and aperiodic Markov chain on finite state space $E = \{1, 2, \dots, K\}$ with transition probability matrix $P = \|p_{ij}\|_{K \times K}$, where $p_{ij} = p(Z_t = j | Z_{t-1} = i)$, $i, j \in \{1, \dots, K\}$, and stationary probability measure $\pi = (\pi_1, \dots, \pi_K)'$. Also given that $Z_t = j$, $H_{t,j}$ (the conditional variance in regime j) is driven by

$$H_{t,j} = a_{0,j} + a_{1,j}y_{t-1}^2(1 - w_{j,t-1}) + a_{2,j}y_{t-1}^2w_{j,t-1} + b_jH_{t-1,j}, \quad (2.2)$$

and each of the weights ($w_{t,j}$) is a logistic function of the past observation as

$$w_{j,t-1} = \frac{1}{1 + \exp(-\gamma_j y_{t-1})} \quad \gamma_j > 0, \quad j = 1, \dots, K, \quad (2.3)$$

which are bounded, $0 < w_{j,t-1} < 1$, and monotonically increasing. The parameter γ_j is called the slope parameter, that explains the speed of transition from negative shocks to positive one: the higher γ_j , the faster the transition. Since $\gamma_j > 0$, the weight function $0 < w_{j,t-1} < 1$ goes to zero when $y_{t-1} \rightarrow -\infty$ and to one when $y_{t-1} \rightarrow +\infty$. As y_{t-1} increases from negative values to positive values the impact of y_{t-1}^2 proceeds proportionally from $a_{1,j}$ to $a_{2,j}$. So $a_{1,j}$ in each state will characterize negative shocks and $a_{2,j}$ positive ones. It refers to the fact that negative and positive shocks have different effects on volatility. When γ_j tending to zero, $w_{j,t-1}$ goes to 1/2 and the MS-STGARCH model tends to MS-GARCH model.

2.1 Forecasting

The conditional variance of MS-STGARCH model is given by

$$\begin{aligned} Var(Y_t | \mathcal{I}_{t-1}) &= \sum_{i=1}^K \alpha_i^{(t)} H_{t,i} = \sum_{i=1}^K \alpha_i^{(t)} (a_{0,i} + a_{1,i}y_{t-1}^2(1 - w_{i,t-1}) \\ &\quad + a_{2,i}y_{t-1}^2w_{i,t-1} + b_iH_{t-1,i}) \end{aligned} \quad (2.4)$$

as $H_{t,i}$ is the conditional variance of i -th state. As using all past observations for forecasting could increase the complexity of the model, we reduce the volume of calculations by proposing a dynamic programming algorithm.

Remark 2.1. The value of $\alpha_j^{(t)}$ is obtained recursively by

$$\alpha_j^{(t)} = \frac{\sum_{m=1}^K f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2})p(Z_{t-1} = m|\mathcal{I}_{t-2})p_{m,j}}{\sum_{m=1}^K f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2})p(Z_{t-1} = m|\mathcal{I}_{t-2})}. \quad (2.5)$$

3 Estimation

For the estimation of parameters, we apply the Bayesian inference, that is extensively used in literature see [Bauwens et al. \(2010\)](#).

Let Y_t be the vector (y_1, \dots, y_t) . In what follows, for the case of two regimes, we are going to estimate the $Z_t = (z_1, \dots, z_t)$ and the parameter vectors $\eta = (\eta_{11}, \eta_{22})$ and $\theta = (\theta_1, \theta_2)$, where $\theta_k = (a_{0k}, a_{1k}, a_{2k}, b_k, \gamma_k)$ for $k = 1, 2$ from the posterior density

$$p(\theta, \eta, Z|Y) \propto p(\theta, \eta)p(Z|\theta, \eta)f(Y|\theta, \eta, Z), \quad (3.1)$$

in which $Y = (y_1, \dots, y_T)$, $Z = (z_1, \dots, z_T)$ and $p(\theta, \eta)$ is the prior of the parameters.

3.1 Sampling z_t

This step is devoted to the sampling of $p(z_t|\eta, \theta, Y_t)$ that is performed by Chib. Suppose $p(z_1|\eta, \theta, Y_0)$ be the stationary distribution of the chain,

$$p(z_t|\eta, \theta, Y_t) \propto f(y_t|\theta, z_t = k, Y_{t-1})p(z_t|\eta, \theta, Y_{t-1}), \quad (3.2)$$

that by the law of total probability $p(z_t|\eta, \theta, Y_{t-1})$ is given by:

$$p(z_t|\eta, \theta, Y_{t-1}) = \sum_{z_{t-1}=1}^K p(z_{t-1}|\eta, \theta, Y_{t-1})\eta_{z_{t-1}z_t}. \quad (3.3)$$

Given the filter probabilities $(p(z_t|\eta, \theta, Y_t))$, we run a backward algorithm, starting from $t = T$ that z_T is derived from $p(z_T|\eta, \theta, Y)$. For $t = T-1, \dots, 0$ the sample is derived from $p(z_t|z_{t+1}, \dots, z_T, \theta, \eta, Y)$, which is obtained by

$$p(z_t|z_{t+1}, \dots, z_T, \theta, \eta, Y) \propto p(z_t|\eta, \theta, Y_t)\eta_{z_t, z_{t+1}}.$$

To derive z_t from $p(z_t|\cdot) = p_{z_t}$ is by sampling from the conditional probabilities $q_j = p(Z_t = j|Z_t \geq j, \cdot)$ which are given by

$$p(Z_t = j|Z_t \geq j, \cdot) = \frac{p_j}{\sum_{l=j}^K q_l}.$$

After generating a uniform (0,1) number U , if $U \leq q_j$ then $z_t = j$, otherwise increase j to $j + 1$ and generate another uniform (0,1) and compare it by q_{j+1} .

3.2 Sampling η

This stage is devoted to sample $\eta = (\eta_{11}, \eta_{22})$ from the posterior probability $p(\eta|\theta, Y_t, Z_t)$ that is independent of Y_t, θ . We consider independent beta prior density for each of η_{11} and η_{22} .

3.3 Sampling θ

To sample from the $p(\theta|Y, Z, \eta)$ we use the Griddy Gibbs algorithm that introduced by Ritter and Tanner. This method is very applicable in researches (see [Bauwens et al. \(2010\)](#)). Given samples at iteration r the Griddy Gibbs at iteration $r + 1$ proceeds as follows:

Algorithm 1. *This algorithm has the following steps:*

- *Step 1 Select a grid of points, such as $a_{0i}^1, a_{0i}^2, \dots, a_{0i}^G$. Using (3.3), evaluate the conditional posterior density function $k(a_{0i}|Z_t, Y_t, \theta_{-a_{0i}})$ over the grid points to obtain the vector $G_k = (k_1, \dots, k_G)$.*
- *Step 2 By a deterministic integration rule using the G points, compute $G_\Phi = (0, \Phi_2, \dots, \Phi_G)$ with*

$$\Phi_j = \int_{a_{0i}^1}^{a_{0i}^j} k(a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t) da_{0i}, \quad i = 2, \dots, G. \quad (3.4)$$
- *Step 3 Simulate $u \sim U(0, \Phi_G)$ and invert $\Phi(a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t)$ by numerical interpolation to obtain a sample $a_{0i}^{(r+1)}$ from $a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t$.*
- *Step 4 Repeat steps 1-3 for other parameters.*

In this section we apply the daily stock market index of *S&P500* from 04/08/2010 to 24/04/2012 (450 observations) for estimation. Figure 1 demonstrates the stock market index, the percentage

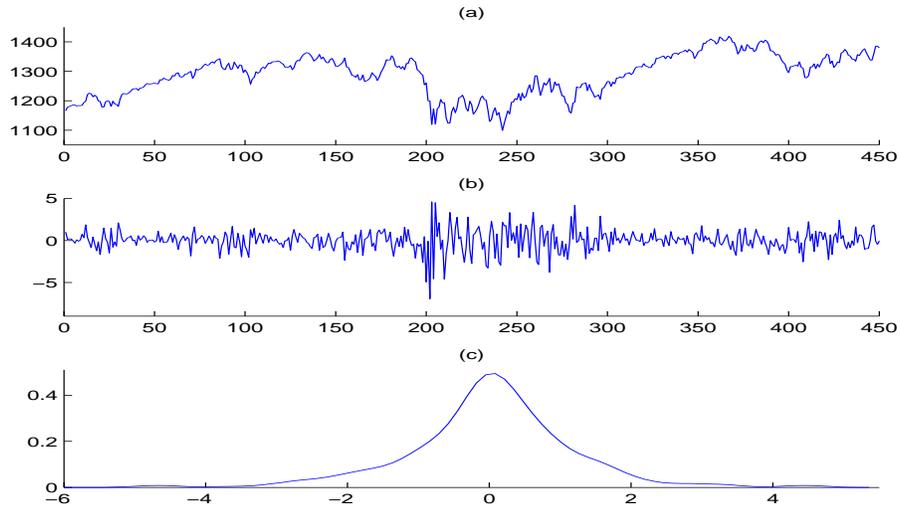


Figure 1: (a): S&P500 from 04/08/2010 to 24/04/2012 stock market index, (b): Percentage daily returns , (c): kernel density of the returns.

returns ¹ and the kernel density of the data set. Using the Bayesian inference, we estimate the parameters of the MS-STGARCH (2.1)-(2.3) and MS-GARCH models (that each model has two regimes). The prior density of each parameter is assumed to be uniform restricted over a finite interval (except for η_{11} and η_{22} , since they are drawn from the beta distribution). Parameters γ_1 and γ_2 in the MS-GARCH model are zero. The estimation results for both models are reported in Table 2 and show that the standard deviations are small enough in most cases. The outcomes of estimating display that the MS-STGARCH has different reactions to positive and negative shocks in both regimes while the MS-GARCH has the similar responses to positive and negative shocks. Since in the MS-STGARCH, $a_{11} > a_{21}$ and according to structure of $w_{j,t-1}$ (2.3) we deduce that the negative shocks have more effect on volatility than the positive shocks in the two regimes. Figure 2 pictures the logistic transition functions for both regimes in MS-STGARCH.

For clarifying the efficiency of the MS-STGARCH model toward the MS-GARCH model, we compare the forecasting volatility ($E(Y_t^2|\mathcal{F}_{t-1})$) of each model with the squared returns .Also for evaluation

¹Percentage return is defined as $r_t = 100 * \log(\frac{P_t}{P_{t-1}})$, where P_t is the index level at time t.

Table 1: Posterior means and standard deviations (*S&P500* daily returns).

	MS-STGARCH		MS-GARCH	
	Mean	Std.dev.	Mean	Std.dev
a_{01}	0.200	0.007	0.199	0.007
a_{11}	0.288	0.008	0.288	0.007
a_{21}	0.103	0.004	0.103	0.005
b_1	0.290	0.006	0.300	.007
γ_1	2.157	0.395	0.000	0.000
a_{02}	0.585	0.054	0.584	0.055
a_{12}	0.673	0.051	0.664	0.046
a_{22}	0.3463	0.039	0.362	0.048
b_2	0.310	0.041	0.311	0.042
γ_2	1.049	0.265	0.000	0.000
η_{11}	0.948	0.029	0.942	0.044
η_{22}	0.9674	0.030	0.968	0.026

of forecasting accuracy of our model toward the MS-GARCH model we compute some measures of performance forecasting ² that are reported in Table 3. The results of this table display that all performance measures of our model are smaller than that of MS-GARCH model.

Table 2: Forecast performance measures

Model	Mean Square Error	Mean Absolute Error	Mean Forecast Error
MS-GARCH	3.211	0.924	-0.178
MS-STGARCH	1.930	0.811	-0.254

Conclusion

In this paper an extension of the MS-GARCH model is introduced where the volatility in each state captures the leverage effect property of financial time series. For the estimation of parameters the Bayesian inference is used by applying the Griddy Gibbs algorithm. Through an empirical example,

²If e_t is the forecast error (the difference between the actual value and the forecasted value), some of the performance measures are as follows:

- Mean absolute error (MAE) = $\frac{1}{n} \sum_{t=1}^n |e_t|$,
- Mean forecast error (MFE) = $\frac{1}{n} \sum_{t=1}^n e_t$,
- Mean squared error (MSE) = $\frac{1}{n} \sum_{t=1}^n e_t^2$.

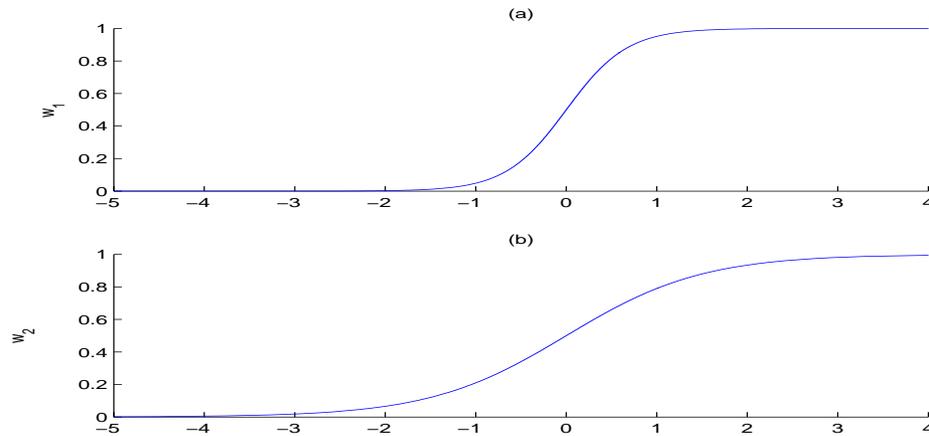


Figure 2: Logistic smooth transition functions for the MS-STGARCH model, (a): first regime (w_1), (b): second regime (w_2).

we show that our model can provide better forecast of volatility than the MS-GARCH.

This model has the potential to be applied for modeling and forecasting conditional volatility of financial time series . For the sake of simplicity it was assumed that the process conditional mean is zero, this assumption could be relaxed by refining the structure of model to allow ARMA structure for conditional mean.

References

- Bauwens, L. Preminger, A. Rombouts, V.K. (2010). Theory and inference for Markov switching GARCH model, *Econometrics journal*, **13**, 218–244.
- Billio, L., Casarin, R. Osuntuyi, A. (2014). Efficient Gibbs sampling for Markov switching GARCH models , *Computational Statistics and Data Analysis*, In Press.
- Engle, R. F.(1990). Discussion: stock volatility and the crash of '87, *Review of Financial Studies*, **3**, 103-106.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, **42**, 27–62.

Haas, M. (2010). Skew normal mixture and Markov-switching GARCH processes, *Studies in Nonlinear Dynamic & Econometrics*, **14**, 1–54.

Haas, M. Paoletta, M.S.(2012). Mixture and Regime-Switching GARCH Models, *Handbook of Volatility Models and Their Applications, First Edition*, John Wiley & Sons, Inc.

Zakoian, J. M. (1994). Threshold heteroskedastic models, *Journal of Economic Dynamic and Control*, **27**, 577–597.



A Sufficient Condition for Absolute Continuity of the Distribution of Lévy Processes

Kasra Alishahi¹, Erfan Salavati²

^{1,2}Department of Mathematical Sciences, Sharif University of Technology.

Abstract: We consider pure jump Lévy processes with discrete and symmetric Lévy measure and study the absolute continuity of their distribution with respect to Lebesgue measure. We prove that if $\eta(r) = \int_{|x|<r} x^2 \nu(dx)$ where ν is the Lévy measure, then $\int_0^1 \frac{r}{\eta(r)} dr < \infty$ is a sufficient condition for absolute continuity. As far as we know, our result is not implied by existing results about absolute continuity of Lévy processes.

Keywords Lévy Processes, Absolute Continuity, Coupling.

Mathematics Subject Classification (2010): 99X99 99X99 99X99.

1 Introduction

Lévy processes constitute an interesting class of stochastic processes which are important from both the theoretical and practical point of view. The problem of absolute continuity of Lévy processes is well studied but no necessary and sufficient condition on the Lévy measure has been found yet. See the references [Sato \(1999\)](#), [Sato \(1982\)](#), [Kallenberg \(1981\)](#), [Orey \(1968\)](#) and [Nourdin and Simon \(2006\)](#).

The cases where the Lévy process has Brownian part or the Lévy measure has a continuous part are easily seen to have absolute continuous distributions. Hence the challenging case is for pure jump Lévy processes with singular Lévy measures.

In this article we consider the special case of real valued Lévy processes with discrete and symmetric Lévy measures of this form:

²Speaker's Name : Kasra Alishahi alishahi@sharif.ir

$$\nu = \sum_{n=1}^{\infty} \delta_{c_n} - \delta_{-c_n}$$

where c_n is a decreasing sequence of real numbers with the condition

$$\sum_{n=1}^{\infty} c_n^2 < \infty.$$

Let X_t be a Lévy process starting at 0 without drift and Brownian part and with Lévy measure ν . We call such a process a Lévy process with characteristics $(0, 0, \nu)$.

Our purpose is to provide a sufficient condition in terms of c_n 's under which the distribution of X_1 is absolutely continuous with respect to Lebesgue measure. We denote the Lebesgue measure on \mathbb{R} by λ .

The main result of this article is Theorem 2.4.

2 Results

We will use the following lemma:

Lemma 2.1. *Let μ be a Borel measure on \mathbb{R} . For $a \in \mathbb{R}$ let μ_a be the push-forward of μ with translation by a . Then $\mu \ll \nu$ if and only if*

$$\lim_{a \rightarrow 0} \|\mu_a - \mu\|_{TV} = 0$$

where $\|\cdot\|_{TV}$ is the total variation norm.

Proof. Assume $\lambda(A) = 0$. Hence for any $x \in \mathbb{R}$, $\lambda(A+x) = 0$. By Fubiny's theorem,

$$\begin{aligned} 0 &= \int_{\mathbb{R}} \lambda(A+x) d\mu(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x+y) d\lambda(y) d\mu(x) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x+y) d\lambda(y) d\mu(x) = \int_{\mathbb{R}} \mu_y(A) d\lambda(y) \quad (2.1) \end{aligned}$$

On the other hand, by hypothesis, the function $y \mapsto \mu_y(A)$ is continuous at $y = 0$ and since it is nonnegative, hence it follows from (2.1) that $\mu(A) = \mu_0(A) = 0$. \square

The main idea of this article is a coupling of Lévy processes. In fact let X_t and Y_t be Lévy processes with characteristic $(0, 0, \nu)$ starting respectively at 0 and a . Without loss of generality we assume $a > 0$. If we denote the distribution of X_1 by μ , then the distribution of Y_1 is μ_a .

By a coupling between X_t and Y_t , we mean a two dimensional process (\bar{X}_t, \bar{Y}_t) where \bar{X}_t has the same distribution as X_t and \bar{Y}_t has the same distribution as Y_t . In the following, we provide a coupling between X_t and Y_t and use it to satisfy the conditions of lemma 2.1.

We denote the left limit of a cadlag process X_t at t by X_{t-} and let $\Delta X_t = X_t - X_{t-}$.

Let $\bar{X}_t = X_t$ and

$$\bar{Y}_t = a + \sum_{s \leq t} \sigma(s) \Delta X_s$$

where

$$\sigma(s) = \begin{cases} 1 & \frac{|X_s - Y_s|}{2} < |\Delta X_s| \\ \xi(s) & \text{otherwise} \end{cases}$$

and $\xi(s)$ are independent random variables with values ± 1 and probability $\frac{1}{2}$ and are independent of X_t .

In other words, whenever \bar{X}_t has a jump greater than half of the distance of \bar{X}_t and \bar{Y}_t , \bar{Y}_t does the same jump and otherwise \bar{Y}_t does a random jump with the same length but in random direction. It is obvious that the distribution of \bar{Y}_t is the same as Y_t and also by method of construction of \bar{Y}_t , we have always $\bar{X}_t \leq \bar{Y}_t$ and $\bar{Y}_t - \bar{X}_t$ is a decreasing process.

Let

$$\tau_a = \inf\{t : \bar{X}_t = \bar{Y}_t\}$$

We state our sufficient condition in terms of the auxiliary function

$$\eta(r) = \int_{|x| < r} x^2 \nu(dx) = 2 \sum_{c_n < r} c_n^2$$

Proposition 2.2. *If $\int_0^1 \frac{r}{\eta(r)} dr < \infty$ then $\lim_{a \rightarrow 0} \mathbb{E}\tau_a = 0$.*

Proof. Let

$$Z_t = \bar{Y}_t - \bar{X}_t$$

Z_t is decreasing and nonnegative.

Lemma 2.3. *By the condition imposed on η , there exists a smooth function $g : [0, \infty) \rightarrow \mathbb{R}$ such that $g''\eta \geq 1$.*

proof of lemma. Define g by,

$$g(x) = \int_x^1 \int_y^1 \frac{1}{\eta(r)} dr dy$$

By Fubiny's theorem,

$$g(x) = \int_x^1 \int_0^r \frac{1}{\eta(r)} dy dr = \int_x^1 \frac{r}{\eta(r)} dr$$

hence g is finite and continuous at 0. On the other hand, we have $g''(x) = \eta(x)$ in continuity points of η and g .

Now by a simple regularization argument we can make g smooth. □

For function g constructed in lemma, we have

$$\begin{aligned} g(Z_{t \wedge \tau_a}) &= g(a) + \sum_{s \leq t \wedge \tau_a} \Delta g(Z_s) \\ &= g(a) + \sum_{s \leq t \wedge \tau_a} g'(Z_{s-}) \Delta Z_s + \frac{1}{2} \sum_{s \leq t \wedge \tau_a} g''(Z_{s-}) (\Delta Z_s)^2 + \sum_{s \leq t \wedge \tau_a} R(Z_{s-}, \Delta Z_s) \end{aligned} \quad (2.2)$$

where $R(Z_{s-}, \Delta Z_s)$ is the Taylor's remainder.

Second term in (2.2) is a martingale, hence

$$\mathbb{E}g(Z_{t \wedge \tau_a}) = g(a) + \frac{1}{2} \mathbb{E} \sum_{s \leq t \wedge \tau_a} g''(Z_{s-}) (\Delta Z_s)^2 + \mathbb{E} \sum_{s \leq t \wedge \tau_a} R(Z_{s-}, \Delta Z_s)$$

For the third term in (2.2) we have

$$\mathbb{E} \sum_{s \leq t \wedge \tau_a} g''(Z_{s-}) (\Delta Z_s)^2 = \mathbb{E} \int_0^{t \wedge \tau_a} g''(Z_s) \eta(Z_s) ds \geq \mathbb{E}(t \wedge \tau_a)$$

For the fourth term in (2.2) we have by Taylor's remainder theorem

$$R(Z_{s-}, \Delta Z_s) \leq C |\Delta Z_s|^3.$$

and since $0 \leq Z_s \leq a$,

$$R(Z_{s-}, \Delta Z_s) \leq Ca |\Delta Z_s|^2$$

which implies

$$\mathbb{E} \sum_{s \leq t \wedge \tau_a} R(Z_{s-}, \Delta Z_s) \leq Ca \mathbb{E} \int_0^{t \wedge \tau_a} \eta(Z_s) ds \leq Ca \eta(a) \mathbb{E}(t \wedge \tau_a)$$

Hence,

$$\frac{1}{2} \mathbb{E}(t \wedge \tau_a) \leq \mathbb{E} g(Z_{t \wedge \tau_a}) - g(a) + a \eta(a) \leq \max_{0 \leq x \leq a} g(x) - g(a) + Ca \eta(a) \mathbb{E}(t \wedge \tau_a)$$

letting $t \rightarrow \infty$ and $a \rightarrow 0$, by continuity of g the statement is proved. \square

Theorem 2.4. *If $\int_0^1 \frac{r}{\eta(r)} dr < \infty$ then the distribution of X_1 is absolutely continuous with respect to Lebesgue measure.*

Proof. If (\bar{X}_t, \bar{Y}_t) is the coupling introduced above,

$$\begin{aligned} \|\mu_{X_1} - \mu_{Y_1}\|_{TV} &= \|\mu_{\bar{X}_1} - \mu_{\bar{Y}_1}\|_{TV} \\ &= \sup_A |\mathbb{P}(\bar{X}_1 \in A) - \mathbb{P}(\bar{Y}_1 \in A)| \\ &\leq \mathbb{P}(\bar{X}_1 \neq \bar{Y}_1) \\ &= \mathbb{P}(\tau_a > 1) \leq \mathbb{E} \tau_a \rightarrow 0 \end{aligned}$$

Hence the conditions of lemma 2.1 are satisfied and we conclude that μ_{X_1} is absolutely continuous with respect to Lebesgue measure. \square

Remark 2.5. *As far as we know, the result of Theorem 2.4 is not implied by existing results about absolute continuity of Lévy processes. See for example, Sato (1999), Proposition 28.3, which provides the sufficient condition $\liminf_{r \rightarrow 0} \frac{\eta(r)}{r^{2-\alpha}} > 0$ for some $\alpha \in (0, 2)$.*

References

Sato, Ken-iti. Lévy processes and infinitely divisible distributions. Cambridge university press, 1999.

Sato, Ken-iti. "Absolute continuity of multivariate distributions of class L." *Journal of Multivariate Analysis* 12.1 (1982): 89-94.

Kallenberg, Olav. "Splitting at backward times in regenerative sets." *The Annals of Probability* (1981): 781-799.

Orey, Steven. "On continuity properties of infinitely divisible distribution functions." *The Annals of Mathematical Statistics* 39.3 (1968): 936-937.

Nourdin, Ivan, and Thomas Simon. "On the absolute continuity of Lvy processes with drift." *The Annals of Probability* (2006): 1035-1051.



An alternative to general record models: dynamic after-selection estimation

Morteza Amini¹, Nader Nematollahi²

¹ Department of Statistics, School of Mathematics, Statistics and computer Science,
College of Science, University of Tehran, P.O. Box 14155-6455, Tehran, Iran

² Department of Statistics, Allameh Tabataba'i University, Tehran, Iran

Abstract: We introduce the problem of estimation of the parameters of a dynamically selected population in an infinite sequence of random variables and provide its application in the statistical inference based on record values from a non-stationary scheme.

Keywords General record models, Selected population, Uniformly minimum variance unbiased estimator.

Mathematics Subject Classification (2010): 60G70 62C15 62F99 62F12 62P12.

1 Introduction

The problem of estimation after selection has received considerable attention by many researches in the past three decades. Interested readers are referred to, for example, [Gibbons et al. \(1977\)](#) for more details.

In this paper, we introduce and develop the problem of estimation of the parameters of a dynamically selected population from a sequence of infinite populations which is not studied in the literature, according to the best of our knowledge. Let X_1, X_2, \dots be a sequence of random variables where X_i is drawn from population Π_i with corresponding cumulative distribution function (cdf) $F_{\theta_i}(\cdot)$ and probability density function (pdf) $f_{\theta_i}(\cdot)$. The traffic volume trend, daily temperatures, sequences of stock quotes, or sequences of estimators of interior water volume in a dam reservoir are examples of such sequences.

²Speaker's Name : Morteza Amini morteza.amini@ut.ac.ir

Suppose we want to estimate the parameter of the population corresponding to the largest value of the sequence X_1, X_2, \dots yet seen, that is $\theta_{[n]} = \theta_{T_n}$, where $T_1 = 1$, with probability one, and for $n > 1$, $T_n = \min\{j; j > T_{n-1}; X_j > X_{T_{n-1}}\}$. The statistics $U_n = X_{T_n}$ is called the n th upper record. The random parameter $\theta_{[n]}$ is the parameter corresponding to the the n th record value. The parameter of the population corresponding to the smallest value yet seen (lower record value) is defined in a similar way. Our aim is to estimate $\theta_{[n]}$. This happens for example, when we want to estimate the largest value of traffic volume or stock quotes yet seen, the temperature of the coldest day or the largest volume of the coming water into the dam reservoir, up to now.

For simplicity, we denote $\theta_{[n]}^U$ by $\theta_{[n]}$ hereafter. We may write

$$\theta_{[n]} = \sum_{j=n}^{\infty} \theta_j I_j(X_1, X_2, \dots), \quad (1.1)$$

where

$$I_j = I_j(X_1, X_2, \dots) = I(\max\{X_k; T_{n-1} + 1 \leq k \leq j - 1\} < X_{T_{n-1}} < X_j), \quad (1.2)$$

where $I(A)$ stands for the indicator of the event A .

This problem is related to the so-called *general record model*. The geometrically increasing populations, the Pfeifer, the linear drift and the F^α record models are some of the generally used record models. The basics of non-stationary schemes for the record values are due to Nevzorov (1985) and Pfeifer (1989), who considered the so-called F^α -scheme, that is the sequences of independent random variables with distribution $F_k(x) = (F(x))^{\theta_k}$, $k = 1, 2, \dots$, where F is a continuous cdf and θ_k 's are positive parameters. Alternative non-stationary schemes include geometrically increasing populations, linear trend and Pfeifer models.

In all the above models, strict assumptions are made on the sequence of parameters $\{\theta_i\}_{i \geq 1}$. For instance, in F^α record model, the sequence of the parameters is assumed to be known or depend on a fixed unknown parameter. In the linear drift model, a linearly increasing population is assumed as the underlying population. However, certain natural phenomena may behave otherwise. For example, an earthquake is produced by a natural phenomenon which has a pivotal parameter that varies based on an unknown model. In order to predict extremely destructive earthquakes, a very important question

is on the value of the parameters which cause a new record in the sequence of earthquakes? This motivates us to study the problem of dynamic after-selection estimation.

The theoretical results of the dynamic after-selection problem, consisting unbiased estimation of the parameters of the model as well as unbiased estimation of the risk of the estimators are presented in Sections 2 and 3. A real data example is considered in section 4 to illustrate the applicability of the results.

2 Minimum variance unbiased estimation

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$, $\mathbf{X} = (X_1, X_2, \dots)$ and $h_{\mathbf{X}}(\boldsymbol{\theta})$ be a random parameter (a function of \mathbf{X} and $\boldsymbol{\theta}$). Suppose that $h_{\mathbf{X}}(\boldsymbol{\theta})$ is estimated by $\delta(\mathbf{X})$. Under the squared error loss (SEL) function $L(h_{\mathbf{X}}(\boldsymbol{\theta}), \delta(\mathbf{X})) = (h_{\mathbf{X}}(\boldsymbol{\theta}) - \delta(\mathbf{X}))^2$, the estimator $\delta(\mathbf{X})$ is said to be risk unbiased for $h_{\mathbf{X}}(\boldsymbol{\theta})$, if it satisfies

$$\boldsymbol{\theta}(\delta(\mathbf{X})) = \boldsymbol{\theta}(h_{\mathbf{X}}(\boldsymbol{\theta})). \quad (2.1)$$

In this section, we use the U-V method of [Robbins \(1988\)](#), to find Uniformly Minimum Variance Unbiased (UMVU) estimator of $\theta_{[n]}$ under the general F^α model, presented below.

The general F^α model: For $X_i, i = 1, 2, \dots$, consider two families of distributions, the first with X_i having the survival function

$$\bar{F}_{\theta_i}(x) = 1 - F_{\theta_i}(x) = (\bar{G}(x))^{\theta_i^{-1}}, \quad (2.2)$$

and the second with X_i having the cdf

$$F_{\theta_i}(x) = (G(x))^{\theta_i^{-1}}, \quad (2.3)$$

in which $G(x)$ is a cdf, free of θ_i , and $\bar{G}(x) = 1 - G(x)$. We assume G to be known. These are called *proportional hazard rate* and *proportional reversed hazard rate* families, or simply F^α models in the context of record values. Some well-known members of the above families are exponential, Rayleigh, beta, Pareto and Burr Type XII distribution.

By making use of U-V method of [Robbins \(1988\)](#) for the family (2.2), we have the following lemma.

Lemma 2.1. Let X_1, X_2, \dots be a sequence of independent random variables with survival function defined in (2.2). Let $u_j(\mathbf{x})$ be a real-valued function such that for $j = 1, 2, \dots$,

$$(i) E_{\theta}[|u_j(\mathbf{X})|] < \infty, \quad \forall \theta$$

$$(ii) \int_{-\infty}^{x_j} u_j(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots) h(t) dt < \infty, \quad \forall x_j > 0,$$

in which $h = g/\bar{G}$ is the hazard function of G and g is the corresponding pdf of G . Then the functions

$$\nu_j(\mathbf{X}) = \int_{-\infty}^{X_j} u_j(X_1, \dots, X_{j-1}, t, X_{j+1}, \dots) h(t) dt, \quad j = 1, 2, \dots,$$

satisfy

$$E_{\theta}[\nu_j(\mathbf{X})] = E_{\theta}[\theta_j u_j(\mathbf{X})], \quad j = 1, 2, \dots.$$

Proof. For one component problem (i.e., a single random variable $X_j, j \geq 1$), let $\nu(x) = \int_{-\infty}^x u(t)h(t) dt$. Then, we have

$$\begin{aligned} \theta_j E(u(X_j)) &= \int_{-\infty}^{+\infty} u(x) [\bar{G}(x)]^{\theta_j - 1} g(x) dx \\ &= \int_{-\infty}^{+\infty} u(x) \bar{F}_{\theta_j}(x) h(x) dx \\ &= \int_{-\infty}^{+\infty} u(x) h(x) \left\{ \int_x^{+\infty} dF_{\theta_j}(y) \right\} dx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^y u(x) h(x) dx dF_{\theta_j}(y) = \int_{-\infty}^{+\infty} \nu(x) dF_{\theta_j}(x). \end{aligned}$$

For the sequence X_1, X_2, \dots , the result follows by a similar calculation. \square

The next result gives the unbiased estimator of $\theta_{[n]}$, under SEL function, for the general family (2.2).

Theorem 2.2. Assume G to be known and let $H = -\log \bar{G}$ be the cumulative hazard function of G . For the general family (2.2), an unbiased estimator of $\theta_{[n]}$, under the SEL function, is

$$V_3(\mathbf{X}) = H(U_n) - H(U_{n-1}). \quad (2.4)$$

Proof. From (1.1), (1.2) and Lemma 2, an unbiased estimator of $\theta_{[n]}$ is given by

$$\begin{aligned} V_3(\mathbf{X}) &= \sum_{j=n}^{\infty} \nu_j(\mathbf{X}) = \sum_{j=n}^{\infty} \int_0^{X_j} h(t) I_j(X_1, X_2, \dots, X_{j-1}, t, X_j, \dots) dt \\ &= \sum_{j=n}^{\infty} \left\{ \int_{U_{n-1}}^{X_j} h(t) dt \right\} \times I(\max\{X_k; X_{T_{n-1}+1} \leq k \leq j-1\} < U_{n-1} < X_j) \\ &= H(U_n) - H(U_{n-1}). \end{aligned}$$

□

Remark 2.3. Similarly, for the family (2.3), an unbiased estimator for $\theta_{[n]}$, under the SEL function, is $V_4(\mathbf{x}) = R(U_n) - R(U_{n-1})$, where $R = \log G$ is the cumulative reversed hazard function of the known cdf G .

Remark 2.4. Note that (X_1, X_2, \dots) is a complete sufficient statistic for $(\theta_1, \theta_2, \dots)$. Hence, the above unbiased estimators of $\theta_{[n]}$ are indeed UMVU estimators of $\theta_{[n]}$.

3 Estimation of the Risks

To compare the UMVU estimator with other estimators, we need to compute the risk function of the proposed estimators.

Under the SEL function, the risk of an estimator V is $R(V, \theta_{[n]}) = E(V^2) + E(\theta_{[n]}^2) - 2E(V\theta_{[n]})$. The UMVU estimators obtained in Section 3 are functions of (U_n, U_{n-1}) . Suppose we want to estimate the risk of an estimator of $\theta_{[n]}$ which depend on \mathbf{X} only through U_n and U_{n-1} , i.e. $V = V(U_n, U_{n-1})$. Then, we have the following result.

Theorem 3.1. For the general family (??), and under the SEL function, an unbiased estimator of the risk of an estimator $V = V(U_n, U_{n-1})$ of $\theta_{[n]}$ is

$$\begin{aligned} W(U_n, U_{n-1}) &= V^2(U_n, U_{n-1}) + \frac{(H(U_n) - H(U_{n-1}))^2}{2} \\ &\quad - 2 \int_{U_{n-1}}^{U_n} h(t) V(t, U_{n-1}) dt. \end{aligned}$$

Proof. From Lemma 2 we have

$$\begin{aligned}
(\theta_{[n]}^2) &= \sum_{j=n}^{\infty} \theta_j^2(I_j((X))) \\
&= \sum_{j=n}^{\infty} \theta_j \left[\int_{-\infty}^{X_j} h(t) I_j(X_1, \dots, X_{j-1}, t, X_{j+1}, \dots) dt \right] \\
&= \sum_{j=n}^{\infty} \left[\int_{-\infty}^{X_j} h(s) \int_{-\infty}^s h(t) I_j(X_1, \dots, X_{j-1}, t, X_{j+1}, \dots) dt ds \right] \\
&= \left[\int_{U_{n-1}}^{U_n} h(s) \int_{U_{n-1}}^s h(t) dt ds \right] \\
&= \left[\frac{H^2(U_n) - H^2(U_{n-1})}{2} - H(U_{n-1})(H(U_n) - H(U_{n-1})) \right] \\
&= \left[\frac{(H(U_n) - H(U_{n-1}))^2}{2} \right].
\end{aligned}$$

Furthermore

$$\begin{aligned}
(\theta_{[n]} V(U_n, U_{n-1})) &= \sum_{j=n}^{\infty} \theta_j (I_j(\mathbf{X}) V(X_j, U_{n-1})) \\
&= \sum_{j=n}^{\infty} \left(\int_0^{X_j} h(t) V(t, U_{n-1}) \times I_j(X_1, \dots, X_{j-1}, t, X_{j+1}, \dots) dt \right) \\
&= \left(\int_{U_{n-1}}^{U_n} h(t) V(t, U_{n-1}) dt \right).
\end{aligned}$$

This completes the proof. □

An immediate corollary of Theorem 3.1 is as follows.

Corollary 3.2. For the general family (2.2) and under the SEL function,

(i) an unbiased estimator of the risk of $V_3 = H(U_n) - H(U_{n-1})$ is

$$W_3(U_n, U_{n-1}) = \frac{1}{2} (H(U_n) - H(U_{n-1}))^2;$$

(ii) the risk of V_3 is

$$R(H(U_n) - H(U_{n-1}), \theta_{[n]}) = (\theta_{[n]}^2).$$

Remark 3.3. The results for the general family (2.3) can be obtained by replacing $H(\cdot)$ with $R(\cdot) = \log G(\cdot)$ in Theorem 3.1 and Corollary 3.2.

Remark 3.4. Since (X_1, X_2, \dots) is a complete sufficient statistic for $(\theta_1, \theta_2, \dots)$, the above unbiased estimators of $(V, \theta_{[n]})$ are indeed, UMVU estimators of $(V, \theta_{[n]})$.

The following result presents the distribution of the unbiased estimator in the family (2.2). The proof is easy and hence omitted.

Lemma 3.5. *In the general family (??), the random variables $H(U_n) - H(U_{n-1})$, $n \geq 2$ are conditionally independent given $T_n = j_n$, $n \geq 2$, each distributed as exponential with mean θ_{j_n} .*

For the general family with the survival function (2.2), we have $(H(X_i)) = \theta_i$, which candidates $H(U_n)$ as a natural estimator of $\theta_{[n]}$.

The following Corollary of Theorem 3.1 states that the UMVUE dominates the natural estimator.

Corollary 3.6. *For the general family (2.2) and under the SEL function, we have*

$$(H(U_n), \theta_{[n]}) > (H(U_n) - H(U_{n-1}), \theta_{[n]}).$$

Proof. It is easy to see that

$$\begin{aligned} & (H(U_n), \theta_{[n]}) - (H(U_n) - H(U_{n-1}), \theta_{[n]}) \\ &= 2(H(U_n)H(U_{n-1})) - (H^2(U_{n-1})) - 2(H(U_{n-1})\theta_{[n]}) \\ &= 2(H(U_n)H(U_{n-1})) - (H^2(U_{n-1})) - 2(H(U_{n-1})(H(U_n) - H(U_{n-1}))) \\ &= (H^2(U_{n-1})) > 0. \end{aligned}$$

And the proof is complete. □

4 Rainfall data: an illustrative example

In this section, we utilize the data set which represents the records of the amount of annual (January 1-December 31) rainfall in inches recorded at Los Angeles Civic Center LACC during the 100-year period from 1890 until 1989, presented by [Arnold et al. \(1998\)](#) [p.180].

A member of the F^α model (Model 2) with survival function as in (2.2), that is the Rayleigh distribution with cdf

$$F(x) = 1 - \exp \left\{ \frac{-(x-4)^{1.9}}{113.23} \right\}, \quad x > 4, \quad (4.1)$$

is well-fitted to the data. The p -value for two-sample Kolmogorov-Smirnov test is 0.3333. Thus, we take $H(x) = (x-4)^{1.9}$, to be the known cumulative hazard rate function of the base distribution $G(x) = 1 - \exp \{ -(x-4)^{1.9} \}$, $x > 4$.

Suppose that the only observations are the sequence of upper record values as follows:

$$\begin{array}{cccc} 12.69 & 12.84 & 18.72 & 21.96 \\ 23.92 & 27.16 & 31.28 & 34.04. \end{array}$$

We have $\hat{\theta}_{[n]} = H(U_n) - H(U_{n-1}) = (U_n - 4)^{1.9} - (U_{n-1} - 4)^{1.9}$ and the unbiased estimator of its risk is $\hat{R}(\hat{\theta}_{[n]}, \theta_{[n]}) = \frac{(H(U_n) - H(U_{n-1}))^2}{2} = \frac{((U_n - 4)^{1.9} - (U_{n-1} - 4)^{1.9})^2}{2}$.

Figure 1 shows the values of $\hat{\theta}_{[n]}$ and their corresponding 3- σ region

$$\left(\max \left\{ 0, \hat{\theta}_{[n]} - 1.5\sqrt{\hat{R}(\hat{\theta}_{[n]}, \theta_{[n]})} \right\}, \hat{\theta}_{[n]} + 1.5\sqrt{\hat{R}(\hat{\theta}_{[n]}, \theta_{[n]})} \right),$$

under stationary and non-stationary hypotheses.

References

- Arnold B. C., Balakrishnan N. and Nagaraja H. N. (1998). *Records*, John Wiley & Sons, New York.
- Gibbons, J. D., Olkin, I. and Sobel, M. (1977). *Selecting and Ordering Populations. A New Statistical Methodology*. New York: John Wiley and Sons.
- Nelsen, R. B. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics. Springer, New York.
- Nevzorov V.B. (1985). On record times and inter-record times for sequences of non-identically distributed random variables. *Zap. Nauchn. Sere. LOML.*, **142**, 109 – 118.
- Nevzorov V. (1995). Asymptotic distributions of records in non-stationary schemes. *Journal of Statistical Planning and Inference*, **45**, 261 – 273.

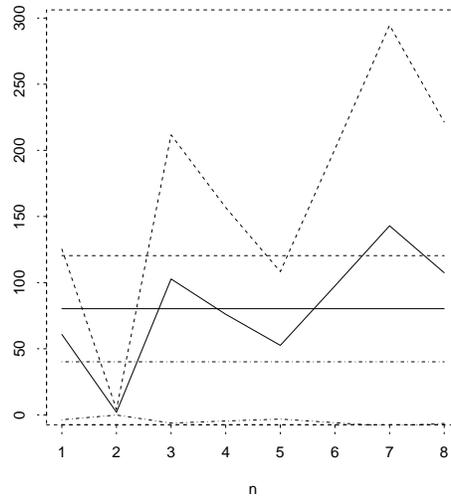


Figure 1: Estimates and $3\text{-}\sigma$ regions of $\theta_{[n]}$, under the stationary and non-stationary assumptions and exponentiality.

Pfeifer D. (1989). Extremal processes, secretary problems and the $1/e$ law, *Journal of Applied Probability*, **27**, 722 – 733.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes.*, Springer-Verlag, New York.

Robbins H. (1988). The U.V methods of estimation. In: Gupta, S.S., Berger, J.O. (Eds.), *Statistical Decision Theory and Related Topics IV*, vol.1. Springer - Verlag, NewYork, pp. 265 - 270.



Truncated unified skew-elliptical distributions with application to order statistics

Mehdi Amiri¹, Ahad Jamalizadeh², Seyyed Roohollah Roozegar³

¹Department of Statistics, Hormozgan University, Bandar Abbas, Iran

²Department of Statistics, Faculty of Mathematics & Computer, Shahid Bahonar University of Kerman, Kerman, Iran

³Department of Mathematics, Yasouj University, Yasouj, Iran

Abstract: We introduce here the truncated version of the unified skew-elliptical (SUE) distributions. By considering special truncations, the joint distribution of consecutive order statistics from an elliptical random vector were derived as the mixture of truncated SUE distributions. This results were applied to determine some measures in the reliability theory.

Keywords: Truncated unified skew-elliptical distribution, Order statistics, Singular unidied skew-normal distribution.

Mathematics Subject Classification (2010): 62E10 62E15 62H10.

1 Introduction

[Arellano-Valle and Azzalini \(2006\)](#) constituted a unified formulation of the skew-normal distribution named unified skew-normal (SUN) distribution. They showed that, this new formulation encompasses all previous proposals, once some redundancies in parametrization are removed. Further they extended the results of the normal framework to the one of elliptical distributions and introduced unified class of skew-elliptical distributions, named unified skew-elliptical (SUE) distribution. Some skew distributions arise from distributions of order statistics. Distributions of order statistics from normal distribution have been discussed by several authors, including [Nagaraja \(1982\)](#) and [Arellano-Valle and Genton](#)

¹Speaker's Name : me.amiry@gmail.com

(2007). [Jamalizadeh and Balakrishnan \(2010\)](#) proved in general that the distributions of order statistics and linear combinations of order statistics from the elliptical distribution are mixtures of SUE distributions.

In this paper, the class of truncated SUE (TSUE) distributions were introduced. The joint distribution of consecutive order statistics arising from an elliptical distributions were derived as a mixture of TSUE distributions. These results were applied to obtain some reliability measures.

2 Preliminaries

The following notations will be used throughout this paper: $\Phi_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the cumulative distribution function (cdf) of $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (n -variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$), $f_{EC_n}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)})$ for the probability density function (pdf) of $EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)})$ (n -variate elliptical distribution with location parameter $\boldsymbol{\mu}$, dispersion matrix $\boldsymbol{\Sigma}$, and density generator function $h^{(n)}$), $F_{EC_n}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)})$ for the cdf of $EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)})$, and simply $\Phi_n(\cdot; \boldsymbol{\Sigma})$ and $F_{EC_n}(\cdot; \boldsymbol{\Sigma}, h^{(n)})$ for the case when $\boldsymbol{\mu} = \mathbf{0}$. Furthermore, for $r \in \mathbb{N}$, let $\mathbf{1}_r$, $\mathbf{0}_r$ and \mathbf{I}_r denote the vector of ones, zeroes and the identity matrix of dimension r , respectively.

2.1 SUE distributions

We give a brief review of the multivariate unified skew-elliptical (SUE) distribution.

Definition 2.1. ([Arellano-Valle and Azzalini \(2006\)](#)) Let \mathbf{U} and \mathbf{V} be jointly distributed as

$$(\mathbf{U}^T, \mathbf{V}^T)^T \sim EC_{m+n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(m+n)}), \quad (2.1)$$

where $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Omega} \end{pmatrix}$. Then let

$$\mathbf{Y} \stackrel{d}{=} \mathbf{V} \mid (\mathbf{U} > \mathbf{0}), \quad (2.2)$$

where $\mathbf{U} > \mathbf{0}$ means that all of the components of \mathbf{U} are bigger than zero. Then, the random vector \mathbf{Y} is said to have the n -variate SUE distribution with set of parameters $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$ and density generator $h^{(m+n)}$, denoted by $\mathbf{Y} \sim SUE_{n,m}(\boldsymbol{\theta}, h^{(m+n)})$.

The pdf of \mathbf{Y} in (2.2), for $\mathbf{y} \in \mathbb{R}^n$, is

$$\frac{f_{EC_n}(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(n)})}{F_{EC_m}(\boldsymbol{\eta}; \boldsymbol{\Gamma}, h^{(m)})} \times F_{EC_m}\left(\boldsymbol{\eta} + \boldsymbol{\Lambda}^T \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \boldsymbol{\Gamma} - \boldsymbol{\Lambda}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda}, h_{w(\mathbf{y})}^{(m)}\right), \quad (2.3)$$

where $w(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})$, $h_{w(\mathbf{y})}^{(m)}(u) = h^{(m+n)}(u + w(\mathbf{y})) / h^{(n)}(u)$ and for $k = m, n$,

$$h^{(k)}(u) = \frac{\pi^{(m+n-k)/2}}{\Gamma((m+n-k)/2)} \int_0^{+\infty} x^{(m+n-k)/2-1} h^{(m+n)}(u+x) dx, \quad u \geq 0.$$

Remark 2.2. If $h^{(m+n)}(u) = (2\pi)^{-(m+n)/2} \exp(-u/2)$ ($u \geq 0$), we obtain the multivariate SUN distribution and for the vector \mathbf{Y} in (2.2), the notation $\mathbf{Y} \sim SUN_{n,m}(\boldsymbol{\theta})$ is used. Also the moment generating function (MGF) of \mathbf{Y} is given by

$$M_{SUN_{n,m}}(\mathbf{s}; \boldsymbol{\theta}) = \frac{\exp\left(\boldsymbol{\xi}^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T \boldsymbol{\Omega} \mathbf{s}\right)}{\Phi_m(\boldsymbol{\eta}; \boldsymbol{\Gamma})} \times \Phi_m(\boldsymbol{\eta} + \boldsymbol{\Lambda}^T \mathbf{s}; \boldsymbol{\Gamma}). \quad (2.4)$$

If $\mathbf{Y} \sim SUN_{n,m}(\boldsymbol{\theta})$ and the covariance matrix $\boldsymbol{\Sigma}$, in (??), is singular, then the distribution of \mathbf{Y} is called singular SUN distribution. This case was introduced by Arellano-Valle and Azzalini (2006) and they considered 3 cases of singularity of $\boldsymbol{\Sigma}$. We mention here the third case.

Remark 2.3. If $\mathbf{Y} \sim SUN_{n,m}(\boldsymbol{\theta})$, $\text{Rank}(\boldsymbol{\Gamma}) = m$ and $\text{Rank}(\boldsymbol{\Omega}) = n$ but $\text{Rank}(\boldsymbol{\Sigma}) < m + n$, then \mathbf{Y} has singular SUN distribution with set of parameters $\boldsymbol{\theta}$, denoted by $\mathbf{Y} \sim SSUN_{n,m}(\boldsymbol{\theta})$. In this case, the MGF of \mathbf{Y} is yet of the form (2.4).

With this setup, in Section 3, the truncated SUE distributions are introduced. In Section 4, the joint distribution of the consecutive order statistics from an elliptical random vector is presented and the results are used to determine some reliability measures.

3 Truncated SUE distributions

The truncated version of SUE distribution is defined as follows.

Definition 3.1. Let the measurable subset $\mathbf{C} \subset \mathbb{R}^n$, $(\mathbf{U}^T, \mathbf{V}^T)^T$ distributed as (2.1) and

$$\mathbf{Y}^{\mathbf{C}} \stackrel{d}{=} \mathbf{V} \mid (\mathbf{U} > \mathbf{0}, \mathbf{V} \in \mathbf{C}). \quad (3.1)$$

Then $\mathbf{Y}^{\mathbf{C}}$ is said to have a truncated SUE distribution with set of parameters $\boldsymbol{\theta}$, density generator function $h^{(m+n)}$ and truncation set \mathbf{C} , denoted by $\mathbf{Y}^{\mathbf{C}} \sim TSUE_{n,m}(\boldsymbol{\theta}, h^{(m+n)}, \mathbf{C})$.

The random vector $\mathbf{Y}^{\mathbf{C}}$ has the pdf, for $\mathbf{y} \in \mathbf{C}$,

$$\frac{f_{EC_n}(\mathbf{y}; \boldsymbol{\xi}, h^{(n)}, \boldsymbol{\Omega})}{P(\mathbf{U} > \mathbf{0}, \mathbf{V} \in \mathbf{C})} \times F_{EC_m}\left(\boldsymbol{\eta} + \boldsymbol{\Lambda}^T \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \boldsymbol{\Gamma} - \boldsymbol{\Lambda}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda}, h_{w(\mathbf{y})}^{(m)}\right). \quad (3.2)$$

Remark 3.2. If $h^{(m+n)}(u) = (2\pi)^{-(m+n)/2} \exp(-u/2)$ ($u \geq 0$), then the random vector $\mathbf{Y}^{\mathbf{C}}$ has truncated n -variate SUN distribution and denoted by $\mathbf{Y}^{\mathbf{C}} \sim TSUN_{n,m}(\boldsymbol{\theta}, \mathbf{C})$.

In the rest, we use the sets $\mathbf{C}_n = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{D}_n \mathbf{y} > \mathbf{0}\}$, $\mathbf{A}_n^t = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{D}_n \mathbf{y} > \mathbf{0}, y_n < t\}$ and $\mathbf{B}_n^t = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{D}_n \mathbf{y} > \mathbf{0}, y_1 > t\}$ where $\mathbf{D}_n \in \mathbb{R}^{(n-1) \times n}$ is difference matrix such that $\mathbf{D}_n \mathbf{Y} = (Y_2 - Y_1, \dots, Y_n - Y_{n-1})$, i.e., the i -th row of \mathbf{D}_n is $\mathbf{e}_{n,i+1}^T - \mathbf{e}_{n,i}^T$, $i = 1, \dots, n-1$, where $\mathbf{e}_{n,1}, \dots, \mathbf{e}_{n,n}$ are the n -dimensional unit basis vectors. Specifically we have $\mathbf{A}_1^t = (-\infty, t)$ and $\mathbf{B}_1^t = (t, +\infty)$.

3.1 Univariate TSUE distributions

By using the truncation sets \mathbf{A}_1^t and \mathbf{B}_1^t , we have

$$Y^{\mathbf{B}_1^t} \stackrel{d}{=} V \mid (\mathbf{U} > \mathbf{0}, V > t), \quad Y^{\mathbf{A}_1^t} \stackrel{d}{=} V \mid (\mathbf{U} > \mathbf{0}, V < t) \quad (3.3)$$

and using (3.2), $Y^{\mathbf{B}_1^t}$ has the pdf, for $y > t$,

$$f_{EC_1}(y - \xi; \omega, h^{(1)}) \times \frac{F_{EC_m}\left(\boldsymbol{\eta} + \boldsymbol{\lambda}(y - \xi)/\omega; \boldsymbol{\Gamma} - \frac{1}{\omega} \boldsymbol{\lambda} \boldsymbol{\lambda}^T, h_{(y-\xi)^2/\omega}^{(m)}\right)}{F_{EC_{m+1}}\left((\boldsymbol{\eta}^T, \xi - t)^T; \boldsymbol{\Sigma}, h^{(m+1)}\right)}, \quad (3.4)$$

where $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\lambda} \\ \boldsymbol{\lambda}^T & \omega \end{bmatrix}$. By noting Remark 2.3 and (3.3), the following remark is derived.

Remark 3.3. If $Y^{\mathbf{B}_1^t} \sim TSUN_{1,m}(\boldsymbol{\theta}, \mathbf{B}_1^t)$ then $Y^{\mathbf{B}_1^t} \sim SSUN_{1,m+1}(\xi, \boldsymbol{\eta}^*, \omega, \boldsymbol{\Sigma}, \boldsymbol{\lambda}^{*T})$ where $\boldsymbol{\theta} = (\xi, \boldsymbol{\eta}, \omega, \boldsymbol{\Gamma}, \boldsymbol{\lambda}^T)$, $\boldsymbol{\eta}^* = (\boldsymbol{\eta}^T, \xi - t)^T$, $\boldsymbol{\lambda}^* = (\boldsymbol{\lambda}^T, \omega)^T$ and $\boldsymbol{\Sigma}$ defined in (3.4).

If $M_{Y^{\mathbf{B}_1^t}}(s)$ denotes the MGF of $Y^{\mathbf{B}_1^t} \sim TSUN_{1,m}(\boldsymbol{\theta}, \mathbf{B}_1^t)$, we have from Remark 3.3 and (2.4),

$$M_{Y^{\mathbf{B}_1^t}}(s) = \exp\left(\xi s + \frac{1}{2} \omega s^2\right) \frac{\Phi_{m+1}\left(\frac{\boldsymbol{\eta} + \boldsymbol{\lambda} s}{\xi - t + \omega s}; \boldsymbol{\Sigma}\right)}{\Phi_{m+1}\left(\frac{\boldsymbol{\eta}}{\xi - t}; \boldsymbol{\Sigma}\right)}, \quad s \in \mathbb{R}. \quad (3.5)$$

3.2 Multivariate TSUE distribution

Let $(\mathbf{U}^T, \mathbf{V}^T)^T$ be distributed as (2.1), then from (2.2) we have

$$\begin{aligned} \mathbf{Y}^{\mathbf{C}_n} &\stackrel{d}{=} \mathbf{V} | (\mathbf{U} > \mathbf{0}, V_1 < \dots < V_n), \\ \mathbf{Y}^{\mathbf{A}_n^t} &\stackrel{d}{=} \mathbf{V} | (\mathbf{U} > \mathbf{0}, V_1 < \dots < V_n < t), \\ \mathbf{Y}^{\mathbf{B}_n^t} &\stackrel{d}{=} \mathbf{V} | (\mathbf{U} > \mathbf{0}, t < V_1 < \dots < V_n). \end{aligned} \quad (3.6)$$

By representations (3.6) we have the following corollary.

Corollary 3.4. *If $\mathbf{Y}^{\mathbf{C}_n} = (Y_1, \dots, Y_n)^T \sim TSUE_{n,m}(\boldsymbol{\theta}, h^{(m+n)}, \mathbf{C}_n)$, then for $t \in \mathbb{R}$,*

- (i) $\mathbf{Y}^{\mathbf{C}_n} | (Y_1 > t) \sim TSUE_{n,m}(\boldsymbol{\theta}, h^{(m+n)}, \mathbf{B}_n^t)$,
- (ii) $\mathbf{Y}^{\mathbf{C}_n} | (Y_n < t) \sim TSUE_{n,m}(\boldsymbol{\theta}, h^{(m+n)}, \mathbf{A}_n^t)$.

In the TSUN case of random vectors (3.6), using Remark 2.3, we have

Remark 3.5. (i) $TSUN_{n,m}(\boldsymbol{\theta}, \mathbf{A}_n^t) \equiv SSUN_{n,m+n}(\boldsymbol{\xi}, (\boldsymbol{\tau}_{(t-\xi_n)}), \boldsymbol{\Omega}, \boldsymbol{\Psi}^A, (\boldsymbol{\Delta}, -\boldsymbol{\Omega}\mathbf{e}_{n,n}))$,

(ii) $TSUN_{n,m}(\boldsymbol{\theta}, \mathbf{B}_n^t) \equiv SSUN_{n,m+n}(\boldsymbol{\xi}, (\boldsymbol{\tau}_{(\xi_1-t)}), \boldsymbol{\Omega}, \boldsymbol{\Psi}^B, (\boldsymbol{\Delta}, \boldsymbol{\Omega}\mathbf{e}_{n,1}))$,

where $\boldsymbol{\tau} = \begin{pmatrix} \boldsymbol{\eta} \\ \mathbf{D}_n \boldsymbol{\xi} \end{pmatrix}$, $\boldsymbol{\Delta} = (\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_n)^T = (\boldsymbol{\Lambda}, \boldsymbol{\Omega}\mathbf{D}_n^T)$, $\boldsymbol{\Psi}^A = \begin{bmatrix} \boldsymbol{\Psi} & -\boldsymbol{\delta}_n \\ -\boldsymbol{\delta}_n^T & \omega_{nn} \end{bmatrix}$ and $\boldsymbol{\Psi}^B = \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_1^T & \omega_{11} \end{bmatrix}$.

Applying Remark 3.5 and (2.4), and then Corollary 3.4 yields the following result.

Corollary 3.6. *If $(Y_1, \dots, Y_n)^T \sim TSUN_{n,m}(\boldsymbol{\theta}, \mathbf{C}_n)$, then we have the following MGFs,*

$$\begin{aligned} M_{Y_1 | (Y_n < t)}(s) &= \exp\left(\xi_1 s + \frac{1}{2} \omega_{11} s^2\right) \frac{\Phi_{m+n}\left(\begin{pmatrix} \boldsymbol{\tau} + \boldsymbol{\delta}_1 s \\ t - \xi_n - \omega_{1n} s \end{pmatrix}; \boldsymbol{\Psi}^A\right)}{\Phi_{m+n}\left(\begin{pmatrix} \boldsymbol{\tau} \\ t - \xi_n \end{pmatrix}; \boldsymbol{\Psi}^A\right)}, \\ M_{Y_n | (Y_1 > t)}(s) &= \exp\left(\xi_n s + \frac{1}{2} \omega_{nn} s^2\right) \frac{\Phi_{m+n}\left(\begin{pmatrix} \boldsymbol{\tau} + \boldsymbol{\delta}_n s \\ \xi_1 - t + \omega_{1n} s \end{pmatrix}; \boldsymbol{\Psi}^B\right)}{\Phi_{m+n}\left(\begin{pmatrix} \boldsymbol{\tau} \\ \xi_1 - t \end{pmatrix}; \boldsymbol{\Psi}^B\right)} \end{aligned} \quad (3.7)$$

where $\boldsymbol{\Psi}^A$ and $\boldsymbol{\Psi}^B$ are defined in Remark 3.5.

4 Application to order statistics

Let $\mathbf{X} = (X_1, \dots, X_n)^T \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)})$, $\mathbf{X}_{(1, \dots, n)} = (X_{(1)}, \dots, X_{(n)})^T$ be the vector of order statistics arising from \mathbf{X} and $\mathbf{X}_{(r, \dots, r+k)} = (X_{(r)}, \dots, X_{(r+k)})^T$ (for positive integers r and k such that $r+k = 2, \dots, n$). For $\mathbf{i} = (i_1, \dots, i_k)$, $\mathbf{X}_{\mathbf{i}} = (X_{i_1}, \dots, X_{i_k})^T$ and $\mathbf{X}_{-\mathbf{i}}$ is obtained from \mathbf{X} by deleting components $\mathbf{X}_{\mathbf{i}}$. Further let diagonal matrix $\mathbf{S}_{j_1, \dots, j_{r-1}} = \text{diag}(s_1, \dots, s_{n-k-1})$, for positive integer values $j_1 < \dots < j_{r-1}$, is such that $s_i = \begin{cases} 1 & \text{for } i = j_1, \dots, j_{r-1} \\ -1 & \text{otherwise} \end{cases}$.

Proposition 4.1. *The cdf of $\mathbf{X}_{(r, \dots, r+k)}$ is the mixture of cdfs of TSUE distributions as*

$$F_{\mathbf{X}_{(r, \dots, r+k)}}(\mathbf{x}) = \sum_{\mathbf{i} \in \mathbf{I}} \sum_{\mathbf{j} \in \mathbf{J}, \mathbf{i} \cap \mathbf{j} = \emptyset} P_{\mathbf{ij}} F_{TSUE_{k+1, n-k-1}}(\mathbf{x}; \boldsymbol{\theta}_{\mathbf{ij}}, h^{(n)}, \mathbf{C}_{k+1}), \quad \mathbf{x} \in \mathbb{R}^{k+1},$$

where $\mathbf{I} = \{(i_1, \dots, i_{k+1}) ; 1 \leq i_d \neq i_l \leq n\}$, $\mathbf{J} = \{(j_1, \dots, j_{r-1}) ; 1 \leq j_1 < \dots < j_{r-1} \leq n-k-1\}$ and parameters $\boldsymbol{\theta}_{\mathbf{ij}} = (\boldsymbol{\mu}_{(\mathbf{i})}, \boldsymbol{\eta}_{\mathbf{ij}}, \boldsymbol{\sigma}_{(\mathbf{i})}, \boldsymbol{\Gamma}_{\mathbf{ij}}, \boldsymbol{\Lambda}_{\mathbf{ij}})$ with components

$$\begin{aligned} \boldsymbol{\eta}_{\mathbf{ij}} &= \mathbf{S}_{\mathbf{j}} \left\{ \mathbf{A}_{r,k} \boldsymbol{\mu}_{(i_1, i_{k+1})} - \boldsymbol{\mu}_{(-\mathbf{i})} \right\}, \quad \boldsymbol{\Lambda}_{\mathbf{ij}}^T = \mathbf{S}_{\mathbf{j}} \left\{ \mathbf{A}_{r,k} \boldsymbol{\sigma}_{(i_1, i_{k+1}), (\mathbf{i})} - \boldsymbol{\sigma}_{(-\mathbf{i}), (\mathbf{i})} \right\}, \\ \boldsymbol{\Gamma}_{\mathbf{ij}} &= \mathbf{S}_{\mathbf{j}} \left\{ \mathbf{A}_{r,k} \boldsymbol{\sigma}_{(i_1, i_{k+1})} \mathbf{A}_{r,k}^T + \boldsymbol{\sigma}_{(-\mathbf{i})} - \mathbf{A}_{r,k} \boldsymbol{\sigma}_{(i_1, i_{k+1}), (-\mathbf{i})} - \boldsymbol{\sigma}_{(i_1, i_{k+1}), (-\mathbf{i})}^T \mathbf{A}_{r,k}^T \right\} \mathbf{S}_{\mathbf{j}}, \end{aligned}$$

where $\mathbf{A}_{r,k} = (\mathbf{p}_{r,k}, \mathbf{q}_{r,k})$, $\boldsymbol{\mu}_{(\mathbf{i})} = E(\mathbf{X}_{\mathbf{i}})$, $\boldsymbol{\sigma}_{(\mathbf{i})} = \text{Var}(\mathbf{X}_{\mathbf{i}})$, $\boldsymbol{\sigma}_{(\mathbf{i}), (\mathbf{j})} = \text{Cov}(\mathbf{X}_{\mathbf{i}}, \mathbf{X}_{\mathbf{j}})$ and

$$\boldsymbol{\Psi}_{\mathbf{ij}} = \begin{pmatrix} \boldsymbol{\Gamma}_{\mathbf{ij}} & \boldsymbol{\Lambda}_{\mathbf{ij}}^T \mathbf{D}_{k+1}^T \\ \mathbf{D}_{k+1} \boldsymbol{\Lambda}_{\mathbf{ij}} & \mathbf{D}_{k+1} \boldsymbol{\sigma}_{(\mathbf{i})} \mathbf{D}_{k+1}^T \end{pmatrix}, \quad P_{\mathbf{ij}} = F_{EC_n} \left(\begin{pmatrix} \boldsymbol{\eta}_{\mathbf{ij}} \\ \mathbf{D}_{k+1} \boldsymbol{\mu}_{(\mathbf{i})} \end{pmatrix}; \boldsymbol{\Psi}_{\mathbf{ij}}, h^{(n)} \right).$$

Proof. Let $\mathbf{B}_{\mathbf{ij}} = \{\max\{\mathbf{X}_{\mathbf{j}}\} < X_{i_1} < \dots < X_{i_{k+1}} < \min\{\mathbf{X}_{-\mathbf{i}, -\mathbf{j}}\}\}$, then we can write

$$\begin{aligned} F_{\mathbf{X}_{(r, \dots, r+k)}}(\mathbf{x}) &= \sum_{\mathbf{i} \in \mathbf{I}} P(\mathbf{X}_{(r, \dots, r+k)} = \mathbf{X}_{\mathbf{i}}) P(\mathbf{X}_{\mathbf{i}} \leq \mathbf{x} | \mathbf{X}_{(r, \dots, r+k)} = \mathbf{X}_{\mathbf{i}}) \\ &= \sum_{\mathbf{i} \in \mathbf{I}} \sum_{\mathbf{j} \in \mathbf{J}, \mathbf{i} \cap \mathbf{j} = \emptyset} P(\mathbf{B}_{\mathbf{ij}}) P(\mathbf{X}_{\mathbf{i}} \leq \mathbf{x} | \mathbf{B}_{\mathbf{ij}}). \end{aligned}$$

Further we can write $\mathbf{B}_{\mathbf{ij}} = \{\mathbf{U}_{\mathbf{ij}} > \mathbf{0}, \mathbf{X}_{\mathbf{i}} \in \mathbf{C}_{k+1}\}$ where $\mathbf{U}_{\mathbf{ij}} = \mathbf{S}_{\mathbf{j}} \{\mathbf{A}_{r,k} \mathbf{X}_{i_1, i_{k+1}} - \mathbf{X}_{-\mathbf{i}}\}$. Then, we have $P(\mathbf{B}_{\mathbf{ij}}) = P_{\mathbf{ij}}$ and by using Definition 3.1, the proof is completed. \square

4.1 Exchangeable Elliptical case

Let $\mathbf{X} = (X_1, \dots, X_n)^T \sim EC_n(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n, h^{(n)})$ and

$$\boldsymbol{\mu}_n = \mu \mathbf{1}_n, \quad \boldsymbol{\Sigma}_n = \sigma^2 \{(1 - \rho) \mathbf{I}_n + \rho \mathbf{1}_n \mathbf{1}_n^T\}, \quad \mu \in \mathbb{R}, \quad \sigma > 0, \quad \frac{-1}{n-1} < \rho < 1. \quad (4.1)$$

Then using Proposition 4.1, after some algebra, we have the following result.

Corollary 4.2. $\mathbf{X}_{(r, \dots, r+k)} \sim TSUE_{k+1, n-k-1}((\boldsymbol{\mu}_{k+1}, \mathbf{0}_{n-k-1}, \boldsymbol{\Sigma}_{k+1}, \boldsymbol{\Gamma}_{r,k}, \boldsymbol{\Lambda}_{r,k}), h^{(n)}, \mathbf{C}_{k+1})$ where

$$\begin{aligned} \boldsymbol{\Gamma}_{r,k} &= \sigma^2 (1 - \rho) \{ \mathbf{I}_{n-k-1} + \text{diag}(\mathbf{1}_{r-1} \mathbf{1}_{r-1}^T, \mathbf{1}_{n-k-r} \mathbf{1}_{n-k-r}^T) \}, \\ \boldsymbol{\Lambda}_{r,k} &= \sigma^2 (1 - \rho) (\mathbf{p}_{r,k}, \mathbf{0}_{(n-k-1)} \mathbf{0}_{k-1}^T, -\mathbf{q}_{r,k})^T, \\ \mathbf{p}_{r,k} &= (\mathbf{1}_{r-1}^T, \mathbf{0}_{n-k-r}^T)^T, \quad \mathbf{q}_{r,k} = (\mathbf{0}_{r-1}^T, \mathbf{1}_{n-k-r}^T)^T. \end{aligned}$$

Further $X_{(r)} | (X_{(r)} > t) \stackrel{d}{=} X_r | (\mathbf{U} > \mathbf{0}, X_r \in \mathbf{B}_1^t)$ where $\mathbf{U} = \mathbf{S}_{1, \dots, r-1}(\mathbf{1}_{n-1} X_r - \mathbf{X}_{-r})$.

Remark 4.3. $X_{(r)} | (X_{(r)} > t) \sim TSUE_{1, n-1}((\mu, \mathbf{0}_{n-1}, \sigma^2, \boldsymbol{\Gamma}_r, \boldsymbol{\lambda}_r^T), h^{(n)}, \mathbf{B}_1^t)$ where $\boldsymbol{\Gamma}_r = \sigma^2 (1 - \rho) \{ \mathbf{I}_{n-1} + (\mathbf{1}_{r-1}^T, -\mathbf{1}_{n-r}^T)^T (\mathbf{1}_{r-1}^T, -\mathbf{1}_{n-r}^T) \}$ and $\boldsymbol{\lambda}_r = \sigma^2 (1 - \rho) (\mathbf{1}_{r-1}^T, -\mathbf{1}_{n-r}^T)^T$.

Using Corollary 4.2 and (3.7), the following result is obtained.

Corollary 4.4. If $\mathbf{X} \sim N_n(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, then we have the following MGFs:

$$\begin{aligned} M_{X_{(r)} | X_{(r+k)} < t}(s) &= \exp\left(\mu s + \frac{1}{2} \sigma^2 s^2\right) \frac{\Phi_n\left(\begin{pmatrix} \boldsymbol{\delta}_{r,k}^{\mathbf{A}} s \\ t - \mu - \rho \sigma^2 s \end{pmatrix}; \boldsymbol{\Psi}_{r,k}^{\mathbf{A}}\right)}{\Phi_n\left(\begin{pmatrix} \mathbf{0}_{n-1} \\ t - \mu \end{pmatrix}; \boldsymbol{\Psi}_{r,k}^{\mathbf{A}}\right)}, \quad s \in \mathbb{R}, \\ M_{X_{(r+k)} | X_{(r)} > t}(s) &= \exp\left(\mu s + \frac{1}{2} \sigma^2 s^2\right) \frac{\Phi_n\left(\begin{pmatrix} \boldsymbol{\delta}_{r,k}^{\mathbf{B}} s \\ \mu - t + \rho \sigma^2 s \end{pmatrix}; \boldsymbol{\Psi}_{r,k}^{\mathbf{B}}\right)}{\Phi_n\left(\begin{pmatrix} \mathbf{0}_{n-1} \\ \mu - t \end{pmatrix}; \boldsymbol{\Psi}_{r,k}^{\mathbf{B}}\right)}, \quad s \in \mathbb{R}, \end{aligned}$$

where $\boldsymbol{\delta}_{r,k}^{\mathbf{A}} = \sigma^2 (1 - \rho) \begin{pmatrix} \mathbf{p}_{r,k} \\ -\mathbf{e}_{k,1} \end{pmatrix}$, $\boldsymbol{\delta}_{r,k}^{\mathbf{B}} = \sigma^2 (1 - \rho) \begin{pmatrix} -\mathbf{q}_{r,k} \\ \mathbf{e}_{k,k} \end{pmatrix}$ and

$$\begin{aligned} \boldsymbol{\Psi}_{r,k}^{\mathbf{A}} &= \begin{pmatrix} \boldsymbol{\Psi}_{r,k} & -\boldsymbol{\delta}_{r,k}^{\mathbf{B}} \\ -(\boldsymbol{\delta}_{r,k}^{\mathbf{B}})^T & \sigma^2 \end{pmatrix}, \quad \boldsymbol{\Psi}_{r,k}^{\mathbf{B}} = \begin{pmatrix} \boldsymbol{\Psi}_{r,k} & \boldsymbol{\delta}_{r,k}^{\mathbf{A}} \\ (\boldsymbol{\delta}_{r,k}^{\mathbf{A}})^T & \sigma^2 \end{pmatrix}, \\ \boldsymbol{\Psi}_{r,k} &= \begin{pmatrix} \boldsymbol{\Gamma}_{r,k} & (\mathbf{D}_{k+1} \boldsymbol{\Lambda}_{r,k})^T \\ \mathbf{D}_{k+1} \boldsymbol{\Lambda}_{r,k} & \sigma^2 (1 - \rho) \mathbf{D}_{k+1} \mathbf{D}_{k+1}^T \end{pmatrix}. \end{aligned}$$

The following corollary is consequence of Remark 4.3 and (3.5).

Corollary 4.5. $M_{X_{(r)}|X_{(r)}>t}(s) = \exp\left(\mu s + \frac{1}{2}\sigma^2 s^2\right) \frac{\Phi_n\left(\left(\begin{smallmatrix} \lambda_r s \\ \mu - t + \sigma^2 s \end{smallmatrix}\right); \Psi_r\right)}{\Phi_n\left(\left(\begin{smallmatrix} \mathbf{0}_{n-1} \\ \mu - t \end{smallmatrix}\right); \Psi_r\right)}$, where Γ_r and λ_r defined in Remark 4.3 and $\Psi_r = \begin{pmatrix} \Gamma_r & \lambda_r \\ \lambda_r^T & \sigma^2 \end{pmatrix}$.

4.2 Some reliability measures

In the reliability theory, there have been defined some measures such as the mean past life function (MPL) and the mean residual life function (MRL). If the life time of a system is the random variable T , then these measures are defined respectively by $M(t) = E(t - T|T < t)$ and $m(t) = E(T - t|T > t)$, at time $t > 0$. A common structure of redundancy is the k-out-of-n systems. A k-out-of-n system consisting of n components operates if and only if at least k out of n components are functioning. Especial cases of k-out-of-n systems are parallel and series systems. We use the log-normal distribution as the life time distribution of the k-out-of-n systems. The log-normal distribution is commonly used to model the lives of units whose failure modes are of a fatigue-stress nature. It has been called the most commonly used life distribution model for many high-technology applications.

Let T_1, \dots, T_n denote the lifetimes of the components of a system which T_i 's have a common log-normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ and the correlation coefficient of $Ln(T_i)$ and $Ln(T_j)$ is ρ , for any $i \neq j$. If $T_{(1)}, \dots, T_{(n)}$ are the order statistics corresponding to T_i 's, then $(T_{(1)}, \dots, T_{(n)})^T \stackrel{d}{=} (e^{X_{(1)}}, \dots, e^{X_{(n)}})^T$ where $(X_1, \dots, X_n)^T \sim N_n(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ are as in (4.1). Using Corollary 4.5, the MRL of the series, parallel and k-out-of-n systems are obtained by replacing $r = 1, n, n - k + 1$, respectively, in the following expression,

$$E(T_{(r)} - t|T_{(r)} > t) = \exp\left(\mu + \frac{1}{2}\sigma^2\right) \frac{\Phi_n\left(\left(\begin{smallmatrix} \lambda_r \\ \mu - \ln(t) + \sigma^2 \end{smallmatrix}\right); \Psi_r\right)}{\Phi_n\left(\left(\begin{smallmatrix} \mathbf{0}_{n-1} \\ \mu - \ln(t) \end{smallmatrix}\right); \Psi_r\right)} - t.$$

A new definition for the MRL of the parallel system is proposed by Asadi and Bayramov (2005) as $m_r(t) = E(T_{(n)} - t|T_{(r)} > t)$. Recently, Asadi (2006) has defined the MPL of components of the

system as $M_r(t) = E(t - T_{(r)} | T_{(n)} \leq t)$. From Corollary 4.4, these measures are as follows

$$m_r(t) = \exp\left(\mu + \frac{1}{2}\sigma^2\right) \frac{\Phi_n\left(\left(\begin{smallmatrix} \delta_{r,n-r}^{\mathbf{B}} \\ \mu - \ln(t) + \rho\sigma^2 \end{smallmatrix}\right); \Psi_{r,n-r}^{\mathbf{B}}\right)}{\Phi_n\left(\left(\begin{smallmatrix} \mathbf{0}_{n-1} \\ \mu - \ln(t) \end{smallmatrix}\right); \Psi_{r,n-r}^{\mathbf{B}}\right)} - t,$$

$$M_r(t) = t - \exp\left(\mu + \frac{1}{2}\sigma^2\right) \frac{\Phi_n\left(\left(\begin{smallmatrix} \delta_{r,n-r}^{\mathbf{A}} \\ \ln(t) - \mu - \rho\sigma^2 \end{smallmatrix}\right); \Psi_{r,n-r}^{\mathbf{A}}\right)}{\Phi_n\left(\left(\begin{smallmatrix} \mathbf{0}_{n-1} \\ \ln(t) - \mu \end{smallmatrix}\right); \Psi_{r,n-r}^{\mathbf{A}}\right)}.$$

References

- Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions, *Scand. J. of Statist.*, **33**, 561-574.
- Arellano-Valle, R. B. and Genton, M. G. (2007). On the exact distribution of linear combinations of order statistics from dependent random variables, *J. Multivar. Anal.*, **98**, 1876-1894.
- Asadi, M. (2006). On the mean past lifetime of the components of a parallel system, *J. of Statistical Plann. and Infer.*, **136**, 1197-1206.
- Asadi, M., Bairamov, I. (2005). A note on the mean residual life function of a parallel system, *Commu. in Statist. Theo. and Meth.*, **34**, 1-12.
- Jamalizadeh, A. and Balakrishnan, N. (2010). Distributions of order statistics and linear combinations of order statistics from an elliptical distribution as mixtures of skew-elliptical distributions, *J. of Multivar. Anal.*, **101**, 1412-1427.
- Nagaraja, H.N. (1982). A note on linear functions of ordered correlated normal random variables. *Biometrika*, **69**, 284-285.



Distribution of a random sum of exponential random variables

A. Asgharzadeh¹, M. Habibi^{1,2}

¹ Department of Statistics, School of Mathematical Sciences, University of Mazandaran,
Post code 47416-1467, Babolsar, Iran

Abstract: Let W_1, \dots, W_n denote the lifetimes of the n components of a standby system. Then $X = \sum_{i=1}^n W_i$ is the lifetime of the standby system. In this paper, we consider the distribution of X , when the lifetimes components are independent exponential random variables and the number of components in the system is random and follows a discrete uniform distribution. The statistical properties of the distribution are studied. Estimation of the unknown parameters are discussed by the maximum likelihood method. The usefulness of the distribution is illustrated by one real data set.

Keywords. Standby system, Exponential distribution, Uniform distribution, Maximum likelihood estimator. **Mathematics Subject Classification (2010):** 62F10, 62F15, 62N05.

1 Introduction

Recently, [Withers et al. \(2011\)](#) considered a compound poisson-gamma variable as a sum of a random sample from a gamma distribution with sample size an independent poisson random variable. [Bakouch et al. \(2014\)](#) discussed the distribution of the random sum of the exponential random variables when the sample size has a zero truncated binomial distribution. [Kozubowski et al. \(2005\)](#) studied the joint distribution of X and N , where N has a geometric distribution and X is the sum of N i.i.d. exponential variables, independent of N . In this paper we consider the distribution of sum of a random sample of exponential distribution when the sample size has a discrete uniform distribution. Suppose N denote a discrete uniform random variable with the probability mass function specified by $P(N = n) = \frac{1}{k}$, $n = 1, 2, \dots, k$, where $k = 1, 2, \dots$ is an unknown parameter. Given N , let $X = \sum_{i=1}^N W_i$, where

²Speaker's Name : M.habibi@stu.umz.ac.ir

W_1, \dots, W_N are independent and identical random variables independent of N , having the exponential distribution with the parameter λ . Then, the marginal probability density function (pdf) of X is

$$f_X(x) = \sum_{n=1}^k f_{X|N=n}(x) P(N=n) = \sum_{n=1}^k \frac{\lambda^n x^{n-1} e^{-\lambda x}}{k\Gamma(n)} = \lambda \frac{\Gamma(k, \lambda x)}{\Gamma(k+1)}, \quad x > 0, \quad \lambda > 0, \quad k = 1, 2, \dots \quad (1.1)$$

where $\Gamma(a, x)$ is the incomplete gamma function defined by $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$. The corresponding cumulative distribution function (cdf) is as follows

$$F(x) = 1 - \frac{1}{\Gamma(k+1)} \{\Gamma(k+1, \lambda x) - \lambda x \Gamma(k, \lambda x)\}, \quad x > 0, \quad \lambda > 0, \quad k = 1, 2, \dots \quad (1.2)$$

We call the distribution in (1.1) an exponential uniform (EU) distribution and use $X \sim EU(k, \lambda)$ to denote the random variable X have the EU distribution with parameters k and λ . For $k = 1$, the EU distribution reduces to the exponential distribution. Note that in (1.1) and (1.2), k is an integer value. But if we consider k a positive real number, (1.1) and (1.2) remain still pdf and cdf, respectively. Therefore, in this paper, we can consider the general case $EU(k, \lambda)$ with $k > 0$ and $\lambda > 0$.

Some motivations for introducing the new distribution (1.1) are as follows. The first motivation is that as discussed before, the new distribution is related to the distribution of the lifetime of a standby system. In a standby system, a component is used until it wears out and is then immediately replaced by another component of the system. Assume a standby system has N components functioning independently at a given time, where N is a random variable with a discrete uniform distribution and the lifetime of each component, W_i , is an exponential random variable with parameter λ . Then the lifetime of the standby system has the $EU(k, \lambda)$ distribution. The second motivation is that as we shall see later, the failure rate function of (1.1) can be constant, decreasing and increasing depending on its parameter k . The third motivation is that for integer k , the new distribution is a mixture of gamma distributions.

2 Shape

In this section, we discuss the shape characteristics of the PDF $f(x)$ of the EU distribution. The behavior of $f(x)$ at $x = 0$ and $x = \infty$, respectively, are given by

$$f(0) = \frac{\lambda}{k}, \quad f(\infty) = 0.$$

The first derivative of $\log[f(x)]$ for the EU distribution is

$$\frac{d \log f(x)}{dx} = \frac{-\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k, \lambda x)}. \quad (2.1)$$

From equation (2.1) we can see that $\frac{d \log f(x)}{dx} < 0$ for all x , so $f(x)$ is monotonically decreasing for all x and it has a mode at $x_0 = 0$. Plots of the shapes of the probability density function (1.1) for selected values of k and $\lambda = 1$ are given in Figure 1. We have the following proposition:

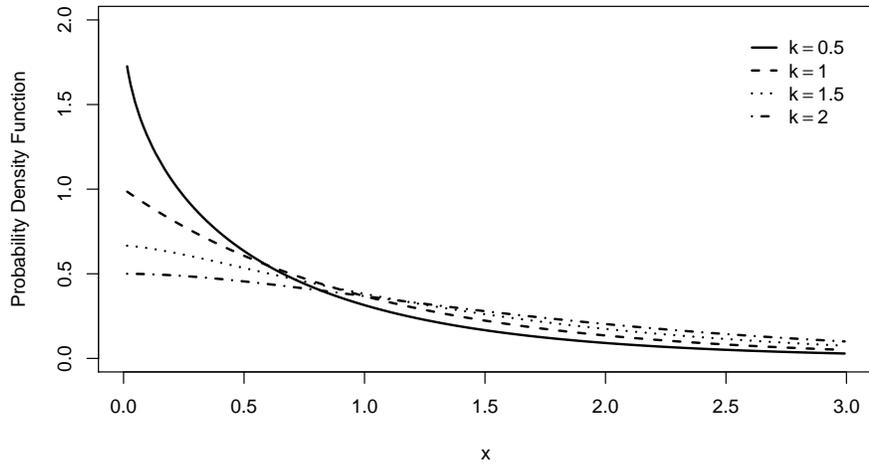


Figure 1: Probability density function of the EU distribution for selected values of k and $\lambda = 1$.

Proposition 2.1. *The pdf of EU distribution is log-convex if $0 < k < 1$, and it is log-concave if $k > 1$.*

The hazard rate function defined by $h(x) = f(x) / [1 - F(x)]$ is an important quantity characterizing life phenomena of a system. For the EU distribution, $h(x)$ takes the form

$$h(x) = \frac{\lambda \Gamma(k, \lambda x)}{\Gamma(k+1, \lambda x) - \lambda x \Gamma(k, \lambda x)}.$$

The behavior of $h(x)$ at $x = 0$ and $x = \infty$, respectively, are given by

$$h(0) = \frac{\lambda}{k}, \quad h(\infty) = \lambda.$$

We can see that the parameter λ represents an upper bound on the hazard rate function which is an important feature of the lifetime models.

Proposition 2.2. *For any $\lambda > 0$, the hazard rate function of EU distribution is decreasing if $0 < k < 1$, and it is increasing if $k > 1$.*

Figure 2 illustrates some of the possible shapes of $h(x)$ for selected values of k and $\lambda = 1$.

3 Statistical measures

In this section, we give some important statistical measures for the EU distribution.

3.1 Quantiles

The p th quantile x_p of the EU distribution defined by $F(x_p) = p$ is the root of the equation

$$x_p = \frac{\Gamma(k+1, \lambda x_p) - (1-p)\Gamma(k+1)}{\lambda \Gamma(k, \lambda x_p)}.$$

Plots of the quartiles of the EU distribution are given in Figure 3 and show an increasing behavior of quartiles. Note that x_p can be used to generate EU random variates.

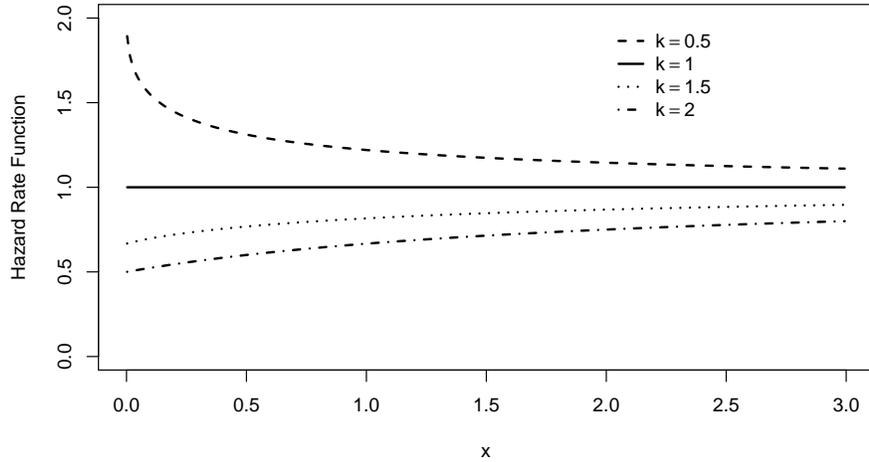


Figure 2: Hazard rate function of the EU distribution for selected values of k and $\lambda = 1$.

3.2 Moment generating function and moments

Let X be a random variable with the probability density function (1.1), the moment generating function of X can be expressed as

$$M_X(s) = E[e^{sX}] = \frac{\lambda}{ks} \left[1 - \left(\frac{\lambda}{\lambda - s} \right)^k \right], \quad s < \lambda.$$

Also, the r th moment (about the origin) of X is given by

$$E(X^r) = \int_0^\infty \frac{\lambda x^r \Gamma(k, \lambda x)}{\Gamma(k+1)} dx = \frac{\Gamma(k+r+1)}{\lambda^r (r+1) \Gamma(k+1)}.$$

In particular, $E(X) = \frac{k+1}{2\lambda}$, $E(X^2) = \frac{(k+1)(k+2)}{3\lambda^2}$ and $Var(X) = \frac{(k+1)(k+5)}{12\lambda^2}$.

Remark 3.1. The central moments of X are $\mu_r = E(X - \mu)^r = \sum_{i=0}^r \binom{r}{i} \mu_i' (-\mu)^{r-i}$, where $\mu = E(X)$. The skewness and kurtosis of X can be obtained using the formulas skewness(X) = μ_3/σ^3 and kurtosis(X) = μ_4/σ^4 , where $\sigma^2 = var(X)$.

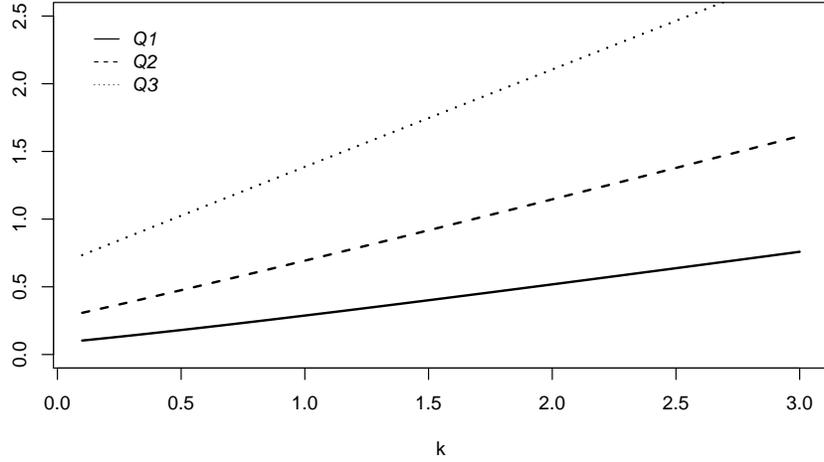


Figure 3: First, second and third quartiles of the EU distribution at $\lambda = 1$.

Figure 4 shows the mean, variance, skewness, and kurtosis of the EU distribution. We see that the mean and variance are increasing as k increases. On the other hand, the skewness and kurtosis are decreasing as k increases.

3.3 Order statistics

Let X_1, X_2, \dots, X_n be a random sample of the $EU(k, \lambda)$ distribution and $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ denote its order statistics. The probability density function and the cumulative distribution function of the i th order statistic are given by

$$F_{i:n}(x) = \sum_{j=i}^n \sum_{l=0}^j \sum_{r=0}^{n-j+l} \binom{n}{j} \frac{(-1)^l}{[\Gamma(k+1)]^{j+l}} (\lambda x)^{rk} (k - \lambda x)^{n+l-j-r} e^{-r\lambda x} [\Gamma(k, \lambda x)]^{n+l-j-r},$$

and

$$f_{i:n}(x) = \frac{\lambda n!}{(i-1)!(n-i)!} \sum_{j=0}^{i-1} \sum_{r=0}^{n+j-i} \sum_{m=0}^{n+j-i-r} \frac{(-1)^{j+m} k^{n+j-i-r-m}}{[\Gamma(k+1)]^{n+j-i+1}} (\lambda x)^{rk+m} e^{-r\lambda x} [\Gamma(k, \lambda x)]^{n+j-i-r},$$

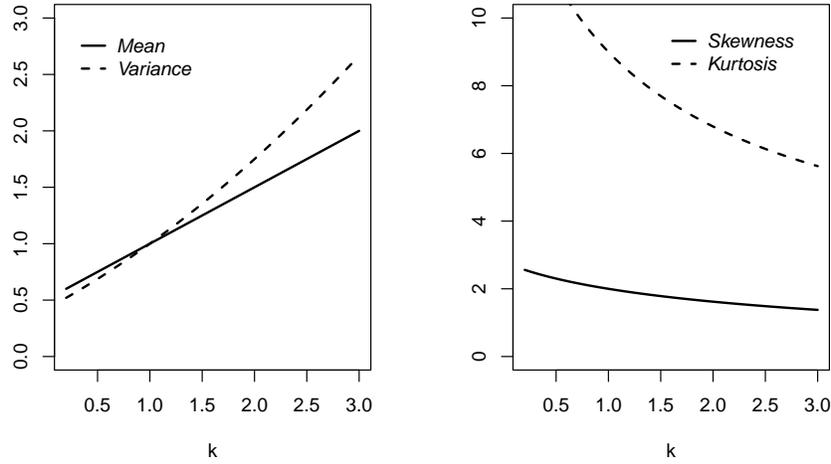


Figure 4: Mean, variance, skewness, and kurtosis of the EU distribution for selected values of k and $\lambda = 1$.

respectively. For integer k , by using the fact $\Gamma(a, x) = \sum_{t=0}^{a-1} \frac{x^t e^{-x}}{t!} \Gamma(a)$, the s th moment of $X_{i:n}$ can be expressed as

$$E(X_{i:n}^s) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{i-1} \sum_{r=0}^{n+j-i} \sum_{m=0}^{n+j-i-r} \sum_{t=0}^{k-1} \frac{(-1)^{j+m} k^{n+j-i-r-m} \Gamma(k) \Gamma(rk+m+s+t+1)}{t! [\Gamma(k+1)]^{n+j-i+1} \lambda^{m+s-1} (r+1)^{rk+m+s+t}}.$$

4 Maximum Likelihood Estimation

Let X_1, X_2, \dots, X_n be a random sample of size n of the $EU(k, \lambda)$ distribution. Let $\boldsymbol{\theta} = (k, \lambda)^T$ be the parameter vector of interest. The log-likelihood function for $\boldsymbol{\theta}$ based on a given random sample is

$$\ln f(\mathbf{x}, \boldsymbol{\theta}) = n \ln(\lambda) - n \ln(\Gamma(k+1)) + \sum_{i=1}^n \ln(\Gamma(k, \lambda x_i)). \tag{4.1}$$

The maximum likelihood estimates of the unknown parameters are obtained by maximizing the log-likelihood function in equation (4.1) with respect to $\boldsymbol{\theta}$. The likelihood equations, which are obtained

from the partial derivatives of $\ln f(\mathbf{X}, \boldsymbol{\theta})$ with respect to the parameters, are

$$\frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial k} = -\frac{n}{k} - n\psi(k) + \sum_{i=1}^n \ln(\lambda x_i) + \sum_{i=1}^n \frac{T(3, k, \lambda x_i)}{\Gamma(k, \lambda x_i)},$$

$$\frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \lambda} = \frac{n}{\lambda} - \lambda^{k-1} \sum_{i=1}^n \frac{x_i^k e^{-\lambda x_i}}{\Gamma(k, \lambda x_i)},$$

where $\psi(\cdot)$ is the digamma function and $T(m, k, x)$ is a special case of the Meijer G-function and defined by

$$T(m, k, x) = G_{m-1, m}^{m, 0} \left(\begin{matrix} 0, \dots, 0 \\ k-1, -1, \dots, -1 \end{matrix} \middle| x \right).$$

The maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{k}, \hat{\lambda})^T$ of $\boldsymbol{\theta} = (k, \lambda)^T$ can be obtained by solving simultaneously the likelihood equations

$$\frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial k} = \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \lambda} = 0.$$

There is no closed-form expression for the maximum likelihood estimator and its computation has to be performed numerically using a nonlinear optimization algorithm.

5 Application

In this section, we fit the EU distribution to a real data set. The data set consists of the number of successive failures of the air conditioning system of each member of a fleet of 13 Boeing 720 jet airplanes. The pooled data, yielding a total of 213 observations, were first analyzed by [Proschan \(1963\)](#) and further discussed in [Gleser \(1989\)](#). We compare the fits of the EU distribution to the fits of the gamma, Weibull, exponentiated exponential, binomial exponential (BE2) and exponential distributions with the respective pdfs $f_G(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, $f_W(x) = \alpha \beta (\beta x)^{\alpha-1} e^{-(\beta x)^\alpha}$, $f_{EE}(x) = \alpha \lambda e^{-\lambda x} (1 - e^{-\lambda x})^{\alpha-1}$, $f_{BE2}(x) = \left(1 + \frac{(\lambda x - 1)\theta}{2 - \theta}\right) \lambda e^{-\lambda x}$ and $f_E(x) = \lambda e^{-\lambda x}$.

We apply formal goodness-of-fit tests in order to verify which distribution fits better this real data set. We consider the Kolmogorov-Smirnov ($K-S$) test statistic and the p -value for ($K-S$) test. Also we use the Cramér-von Mises (W) Anderson-Darling (A), AIC and $AICC$ statistics. In general, the

smaller values of the $K-S$, W , A , AIC and $AICC$ statistics, and the larger value of the p -value indicate the better fit to the data. The parameter estimates and the goodness-of-fit statistics for this data set are given in Table 1. Table 1 shows that the proposed EU model fits the jet airplane data better than the other models according to the statistics $K-S$, W , A , AIC and $AICC$.

Table 1: Estimates, p -value, $K-S$, W , A , AIC and $AICC$ statistics.

Distribution	Estimates	p -value	$K-S$	W	A	AIC	$AICC$
Exponential	$\hat{\lambda} = 0.011$	0.203	0.073	0.165	1.018	2359.532	2357.551
BE2	$\hat{\lambda} = 0.011, \hat{\theta} = 0.018$	0.215	0.072	0.165	1.019	2361.540	2358.596
EE	$\hat{\alpha} = 0.927, \hat{\lambda} = 0.010$	0.328	0.064	0.167	1.031	2360.804	2357.860
Gamma	$\hat{\alpha} = 0.922, \hat{\beta} = 0.009$	0.365	0.062	0.163	1.012	2360.582	2357.638
Weibull	$\hat{\alpha} = 0.924, \hat{\beta} = 0.011$	0.600	0.052	0.138	0.859	2359.170	2356.226
EU	$\hat{k} = 0.537, \hat{\lambda} = 0.008$	0.683	0.049	0.121	0.758	2358.036	2355.092

References

- Aarset, M. V., (1987). How to identify bathtub hazard rate, *IEEE Transactions on Reliability*, **36**, 106–108.
- Bakouch, H. S., Aghababaei Jazi, M., Nadarajah, S., Dolati, A., Roozegar, R., (2014). A lifetime model with increasing failure rate, *Applied Mathematical Modelling*, **38**, 5392–5406.
- Gleser, L. J., (1989). The gamma distribution as a mixture of exponential distributions, *Amer. Statist*, **43**, 115-117.
- Kozubowski, T. J., Panorska, A. K., (2005). A mixed bivariate distribution with exponential and geometric marginals, *Journal of Statistical Planning and Inference*, **134**, 501–520.
- Proschan, F., (1963). Theoretical explanation of observed decreasing failure rate, *Technometrics*, **5**, 375-383.
- Withers, C., Nadarajah, S., (2011). On the compound Poisson-gamma distribution, *Kybernetika*, **47**, 15-37.



Ordering Results for Aggregate Claim Amounts from Two Heterogeneous Marshall-Olkin Extended Weibull Portfolios and their Applications in Insurance Analysis

Ghobad Barmalzan¹, Abedin Haidari, Amir T. Payandeh Najafabadi

Department of Mathematical Sciences, Shahid Beheshti University, G.C. Evin, Tehran,
Iran

Abstract: In this work, we discuss the stochastic comparison of two classical surplus processes in a one-year insurance period. Under the Marshall-Olkin extended Weibull random aggregate claim amounts, we establish some new sufficient conditions for the comparison of aggregate claim amounts in the sense of the usual stochastic order (which implies stop-loss order).

Keywords: Usual Stochastic Order, Stop-Loss Order, Multivariate Chain Majorization, Aggregate Claim Amounts.

Mathematics Subject Classification (2010): 62N05.

1 Introduction

Consider the classical surplus process $U(t)$ given by

$$U(t) = u + ct - \sum_{i=1}^{N(t)} Z_i,$$

where $u = U(0)$, c , Z_i and $N(t)$ denote an initial surplus, constant premium, random claim and a given counting process, respectively. Now consider a situation when there are n policyholders in the given portfolio in an one-year insurance period. Under this assumption, the above classical surplus

¹Ghobad Barmalzan : gh-barmalzan@sbu.ac.ir

process $U(t)$ for $t = 1$ can be restated as

$$U(1) = u + c - \sum_{i=1}^n I_{p_i} X_{\lambda_i}, \quad (1.1)$$

where the random variable X_{λ_i} denotes the total of random claims that can be made in an insurance period and I_{p_i} denotes a Bernoulli random variable associated with X_{λ_i} defined as follows: $I_{p_i} = 1$ whenever the i^{th} policyholder makes random claim X_{λ_i} and $I_{p_i} = 0$ whenever he/she does not make a claim. This sort of restating the classical surplus process $U(t)$ as the surplus process in (1.1) can be made for almost all insurance contracts.

In this work, we then present some stochastic comparisons between two classical surplus processes that can be restated as in (1.1). Since u and c are two constant values, to study any stochastic comparison, we just have to consider only $\sum_{i=1}^n I_{p_i} X_{\lambda_i}$. Indeed, the random variable $S_n(\boldsymbol{\lambda}, \mathbf{p}) = \sum_{i=1}^n I_{p_i} X_{\lambda_i}$ is of interest in various fields of probability and statistics. In particular, in actuarial science, it corresponds to the aggregate claim amount in a portfolio of risks.

The problem of comparing the numbers of claims and aggregate claim amounts with respect to some well-known stochastic orders is of interest on both theoretical and practical grounds. In this direction, [Karlin, S. and Novikoff \(1963\)](#) showed that more dispersion of the p_i 's, according to the vector majorization, implies more dispersion of the total number of claims with respect to the convex order. Specifically, they proved that

$$(p_1, \dots, p_n) \succeq^m (p_1^*, \dots, p_n^*) \implies \sum_{i=1}^n I_{p_i} \leq_{cx} \sum_{i=1}^n I_{p_i^*}, \quad (1.2)$$

where $I_{p_1^*}, \dots, I_{p_n^*}$ are independent Bernoulli random variables, independent of the X_{λ_i} 's, with $E(I_{p_i^*}) = p_i^*$, $i = 1, \dots, n$. [Ma \(2008\)](#) then extended the result in (1.2) to the case of aggregate claim amounts with the same amount of claims and different claims, and showed that if $X_{\lambda_1}, \dots, X_{\lambda_n}$ are non-negative exchangeable random variables, then

$$(p_1, \dots, p_n) \succeq^m (p_1^*, \dots, p_n^*) \implies S_n(\boldsymbol{\lambda}, \mathbf{p}) \leq_{cx} S_n(\boldsymbol{\lambda}, \mathbf{p}^*) \quad (1.3)$$

$$(h(p_1), \dots, h(p_n)) \succeq^m (h(p_1^*), \dots, h(p_n^*)) \implies S_n(\boldsymbol{\lambda}, \mathbf{p}) \geq_{st} S_n(\boldsymbol{\lambda}, \mathbf{p}^*), \quad (1.4)$$

where $h(p) = -\log p$ or $(1-p)/p$. For the case when $(h(p_1), \dots, h(p_n)) \in D_n^+$ and $(h(p_1^*), \dots, h(p_n^*)) \in D_n^+$, where $D_n^+ = \{(x_1, \dots, x_n) : x_1 \geq \dots \geq x_n \geq 0\}$, Ma (2008) also proved that (1.4) holds if $X_{\lambda_1}, \dots, X_{\lambda_n}$ are independent non-negative random variables such that $X_{\lambda_1} \leq_{st} \dots \leq_{st} X_{\lambda_n}$. This was followed up by Frostig (2001) and Hu and Ruan (2004) who established some sufficient conditions for comparing the aggregate claim amounts with respect to the symmetric supermodular, and multivariate usual and symmetric stochastic orders (see Shaked and Shanthikumar (2007)), respectively. Demuit and Frostig (2006) studied the effect of an increase in heterogeneity (in the sense of chain majorization) to the individual model of risk theory according to increasing convex order. Khaledi and Ahmadi (2008) discussed stochastic comparison of two aggregate claims corresponding to two possible different individual risk models in the sense of usual stochastic order.

Recently, Barmalzan et al. (2015) presented a complete version of the results of Khaledi and Ahmadi (2008) to the more general case.

The rest of this paper is organized as follows. Section 2 presents some basic concepts that will be used in the subsequent developments. Section 3 deals with stochastic orderings between order statistics arising from two sets of independent heterogeneous MOEW random variables in terms of the vector majorization between the scale parameters. In Section 4, we use the MOEW distribution as the claim amount distribution and establish some new sufficient conditions for the comparison of aggregate claim amounts in the sense of the usual stochastic order (which implies stop-loss order).

2 Preliminaries

Definition 2.1. Suppose X and Y are two non-negative continuous random variables with distribution functions $F(t) = P(X \leq t)$ and $G(t) = P(Y \leq t)$, and survival functions $\bar{F}(t) = 1 - F(t)$ and $\bar{G}(t) = 1 - G(t)$, respectively.

- (i) X is said to be larger than Y in the usual stochastic order (denoted by $X \geq_{st} Y$) if $\mathbb{E}(\phi(X)) \geq \mathbb{E}(\phi(Y))$ for all increasing functions $\phi : \mathbb{R} \rightarrow \mathbb{R}$ when the involved expectations exist;
- (ii) X is said to be larger than Y in the convex order (denoted by $X \geq_{cx} Y$) if $\mathbb{E}(\phi(X)) \geq \mathbb{E}(\phi(Y))$ for all convex functions $\phi : \mathbb{R} \rightarrow \mathbb{R}$ when the involved expectations exist;

(iii) X is said to be larger than Y in the stop-loss order, or equivalently the increasing convex order (denoted by $X \geq_{sl} Y$), if $\pi_X(d) = \mathbb{E}[(X - d)_+] \geq \pi_Y(d) = \mathbb{E}[(Y - d)_+]$ for all $d > 0$ when the involved expectations exist.

Both the convex and usual stochastic orders imply the stop-loss order. Stop-loss order represents the common preferences of all risk averse decision makers. [Kaas et al. \(1994\)](#) and [Kaas and Hesselager \(1995\)](#) all have studied the higher degree stop-loss transform and order.

Theorem 2.2. ([Shaked and Shanthikumar \(2007\)](#)) Let $g(\cdot)$ be an increasing (decreasing) real-valued function. Then, $X \geq_{st} Y$ implies $g(X) \geq_{st} (\leq_{st})g(Y)$.

Theorem 2.3. ([Shaked and Shanthikumar \(2007\)](#)) Two random variables X and Y satisfy $X \geq_{st} Y$ if and only if there exist two random variables \tilde{X} and \tilde{Y} such that $X \stackrel{st}{=} \tilde{X}$ and $Y \stackrel{st}{=} \tilde{Y}$ and $P(\tilde{X} > \tilde{Y}) = 1$, where $\stackrel{st}{=}$ means the same distribution on both sides of the equality.

Definition 2.4. Suppose $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$ are two random vectors. Then, \mathbf{X} is said to be larger than \mathbf{Y} in the usual multivariate stochastic order (denoted by $\mathbf{X} \geq_{st} \mathbf{Y}$) if $\mathbb{E}(\phi(\mathbf{X})) \geq \mathbb{E}(\phi(\mathbf{Y}))$ for all increasing functions $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ when the involved expectations exist.

The multivariate stochastic ordering implies component-wise usual stochastic ordering. Interested readers may refer to [Shaked and Shanthikumar \(2007\)](#) for comprehensive discussions on univariate and multivariate stochastic orders.

Definition 2.5. For two vectors $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$, let $\{a_{(1)}, \dots, a_{(n)}\}$ and $\{b_{(1)}, \dots, b_{(n)}\}$ denote the increasing arrangements of their components, respectively. Then, the vector \mathbf{a} is said to majorize the vector \mathbf{b} (denoted by $\mathbf{a} \stackrel{m}{\succeq} \mathbf{b}$) if $\sum_{j=1}^i a_{(j)} \leq \sum_{j=1}^i b_{(j)}$ for $i = 1, \dots, n-1$, and $\sum_{j=1}^n a_{(j)} = \sum_{j=1}^n b_{(j)}$.

A square matrix Π is said to be a permutation matrix if each row and column has a single unit, and all other entries to be zero. There are $n!$ such matrices of size $n \times n$, each of which is obtained by interchanging rows (or columns) of the identity matrix I . The matrix of a T-transform (T-transformation) has the form $T = wI + (1 - w)\Pi$, where $0 \leq w \leq 1$ and Π is a permutation matrix (see Chapter 2 of [Marshall et al. \(2011\)](#)).

Definition 2.6. Suppose $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are $m \times n$ matrices, with a_1^R, \dots, a_m^R and b_1^R, \dots, b_m^R as their rows, respectively. Then:

- (i) A is said to row majorize B (denoted by $A \succ^{row} B$) if $a_i^R \succeq b_i^R$ for $i = 1, \dots, m$;
- (ii) A is said to chain majorize B (denoted by $A \succ\!\succ B$) if there exists a finite set of $n \times n$ T_i -transform matrices, $i = 1, \dots, k$, such that $B = AT_1T_2 \dots T_k$.

For an elaborate discussion on the theory of vector and matrix majorizations and their applications, we refer the readers to [Marshall et al. \(2011\)](#).

3 Orders Between Order Statistics from Marshall-Olkin Extended Weibull Distributions

[Marshall and Olkin \(1997\)](#) originally proposed a new family of distributions by adding a shape parameter to a specified distribution. Specifically, suppose H is a baseline distribution function with support \mathcal{R}^+ and corresponding survival function \bar{H} . Then, they introduced the distribution

$$F(t; \alpha) = \frac{H(t)}{1 - \bar{\alpha} \bar{H}(t)}, \quad t, \alpha \in \mathcal{R}^+, \bar{\alpha} = 1 - \alpha. \quad (3.1)$$

Let us now use the Weibull distribution with $\bar{H}(t) = e^{-(\lambda t)^\beta}$ in (3.1). We then have

$$F(t; \alpha, \lambda, \beta) = \frac{1 - e^{-(\lambda t)^\beta}}{1 - \bar{\alpha} e^{-(\lambda t)^\beta}}, \quad t, \alpha, \lambda \in \mathcal{R}^+, \bar{\alpha} = 1 - \alpha, \quad (3.2)$$

The family of distributions in (3.2) is called Marshall-Olkin extended Weibull (MOEW) distribution with shape parameters α, β and scale parameter λ (denoted by $MOEW(\alpha, \lambda, \beta)$). For additional discussion on the MOEW distribution and its applications, we refer the readers to [Ghitany et al. \(2005\)](#) and [Coriderio and Lemonte \(2013\)](#).

Theorem 3.1. ([Hu \(1995\)](#)) Suppose $X_{\lambda_1}, \dots, X_{\lambda_n}$ are independent non-negative random variables with $X_{\lambda_i} \sim F(\lambda_i x)$, $i = 1, \dots, n$. Assume that F is an absolutely continuous distribution function with hazard rate function r . If $r(x)$ and $xr(x)$ are decreasing and increasing in $x \in \mathcal{R}^+$, respectively, then

$$(\lambda_1, \dots, \lambda_n) \stackrel{m}{\succ} (\lambda_1^*, \dots, \lambda_n^*) \implies (X_{1:n}, \dots, X_{n:n}) \geq_{st} (X_{1:n}^*, \dots, X_{n:n}^*).$$

Lemma 3.2. *Suppose $X \sim MOEW(\alpha, 1, \beta)$ with hazard rate r . Then,*

- (i) $r(x)$ is decreasing in $x \in \mathcal{R}^+$ for any $0 < \alpha \leq 1$ and $0 < \beta \leq 1$.
- (ii) $xr(x)$ is increasing in $x \in \mathcal{R}^+$ for any $\alpha > 0$.

Proof. (i) The hazard rate function of X is $r(x) = \beta x^{\beta-1}/1 - \bar{\alpha}e^{-x^\beta}$. Taking derivative of $r(x)$ with respect to x , it readily follows that

$$(r(x))' = \beta(\beta - 1)x^{\beta-2}(1 - \bar{\alpha}e^{-x^\beta}) - \beta^2\bar{\alpha}x^{2\beta-2}e^{-x^\beta}.$$

Now, from the assumption $0 < \alpha \leq 1$ and $0 < \beta \leq 1$, the desired result follows.

(ii) From (3.1), it is easy to observe that $xr(x) = \beta x^\beta/1 - \bar{\alpha}e^{-x^\beta}$. Then, we obtain

$$\begin{aligned} (xr(x))' &\stackrel{sgn}{=} 1 - \bar{\alpha}e^{-x^\beta} - \bar{\alpha}x^\beta e^{-x^\beta} \\ &= m(x), \text{ say,} \end{aligned}$$

where $a \stackrel{sgn}{=} b$ means that a and b have the same sign. It can be readily seen that $m'(x) = \beta\bar{\alpha}x^{2\beta-1}e^{-x^\beta}$, which for $0 < \alpha \leq 1$ ($\alpha \geq 1$) is non-negative (non-positive). Therefore, for $0 < \alpha \leq 1$ ($\alpha \geq 1$), we have $m(x) \geq m(0) = 1 - \alpha$ ($m(x) \geq \lim_{x \rightarrow \infty} m(x) = 1$). Thus, for any $\alpha > 0$, we see that $m(x) > 0$, which completes the proof of the lemma. \square

From Theorem 3.1 and Lemma 3.2, we immediately obtain the following theorem.

Theorem 3.3. *Under the assumptions of Theorem 3.1, if $X_{\lambda_i} \sim MOEW(\alpha, \lambda_i, \beta)$, then for $0 < \alpha \leq 1$ and $0 < \beta \leq 1$,*

$$(\lambda_1, \dots, \lambda_n) \stackrel{m}{\succ} (\lambda_1^*, \dots, \lambda_n^*) \implies (X_{1:n}, \dots, X_{n:n}) \geq_{st} (X_{1:n}^*, \dots, X_{n:n}^*).$$

The following corollary is a direct consequence of Theorem 3.3.

Corollary 3.4. *Under the assumptions of Theorem 3.3, for $0 < \alpha \leq 1$ and $0 < \beta \leq 1$,*

$$(\lambda_1, \dots, \lambda_n) \stackrel{m}{\succ} (\lambda_1^*, \dots, \lambda_n^*) \implies \sum_{i=1}^n X_{\lambda_i} \geq_{st} (\geq_{sl}) \sum_{i=1}^n X_{\lambda_i^*}.$$

4 Use of MOEW distribution as the claim amount distribution

Theorem 4.1. (*Khaledi and Ahmadi (2008)*) Suppose $X_{\lambda_1}, \dots, X_{\lambda_n}$ are independent non-negative random variables with X_{λ_i} having survival function $\bar{F}(\cdot; \lambda_i)$, where $\lambda_i > 0$ for $i = 1, \dots, n$, and that I_{p_1}, \dots, I_{p_n} are independent Bernoulli random variables, independent of the X_{λ_i} 's, with $E(I_{p_i}) = p_i$, $i = 1, \dots, n$. In addition, suppose the following two conditions hold:

- (i) $\bar{F}(\cdot; \lambda_i)$ is a decreasing convex function with respect to λ_i , $i = 1, \dots, n$;
- (ii) The survival function of $\sum_{i=1}^n X_{\lambda_i}$ is Schur-convex in $\boldsymbol{\lambda}$.

Then, for $(\boldsymbol{\lambda}, \mathbf{h}(\mathbf{p})) \in \mathcal{S}_n$ and $(\boldsymbol{\lambda}^*, \mathbf{h}(\mathbf{p}^*)) \in \mathcal{S}_n$, the survival function of $S_n(\boldsymbol{\lambda}, \mathbf{p})$ is Schur-convex in $(\boldsymbol{\lambda}, \mathbf{h}(\mathbf{p}))$ with respect to multivariate chain majorization, where $h(p) = -\log p$ or $h(p) = (1-p)/p$.

It needs to be mentioned here that there is a problem in the above \mathcal{S}_n as introduced by [Khaledi and Ahmadi \(2008\)](#) since it is not permutation invariant; see [Barmalzan et al. \(2015\)](#).

In the following theorem, we show that [Theorem 4.1](#) also holds when the set \mathcal{S}_n is replaced by \mathcal{U}_n . Since the proof of this result is quite similar to that of [Theorem 4.1](#) given by [Khaledi and Ahmadi \(2008\)](#), it is omitted here for the sake of brevity.

Theorem 4.2. Under the assumptions of [Theorem 4.1](#), for $(\boldsymbol{\lambda}, \mathbf{h}(\mathbf{p})) \in \mathcal{U}_n$, the survival function of $S_n(\boldsymbol{\lambda}, \mathbf{p})$ is Schur-convex in $(\boldsymbol{\lambda}, \mathbf{h}(\mathbf{p}))$ with respect to multivariate chain majorization, where $h(p) = -\log p$ or $h(p) = (1-p)/p$.

It is of interest to note that $\mathcal{S}_n \subset \mathcal{U}_n$ and that this implication is strict; that is, there exist some matrices in \mathcal{U}_n which are not in \mathcal{S}_n . Thus, based on [Theorem 4.2](#), we can compare more aggregate claim amounts than those based on [Theorem 4.1](#). The following example illustrates this point.

Next, we prove that the MOEW distribution can be used as the claim amount distribution.

Theorem 4.3. Suppose $X_{\lambda_1}, \dots, X_{\lambda_n}$ are independent random variables with $X_{\lambda_i} \sim \text{MOEW}(\alpha, \lambda_i, \beta)$, $i = 1, \dots, n$, and that I_{p_1}, \dots, I_{p_n} are independent Bernoulli random variables, independent of the X_{λ_i} 's, with $E(I_{p_i}) = p_i$, $i = 1, \dots, n$. Then, for $0 < \alpha \leq 1$, $0 < \beta \leq 1$ and $(\boldsymbol{\lambda}, \mathbf{h}(\mathbf{p})) \in \mathcal{U}_n$, the survival function of $S_n(\boldsymbol{\lambda}, \mathbf{p})$ is Schur-convex in $(\boldsymbol{\lambda}, \mathbf{h}(\mathbf{p}))$ with respect to multivariate chain majorization, where $h(p) = -\log p$ or $h(p) = (1-p)/p$.

Proof. From (3.1), we have

$$\bar{F}(t; \lambda) = \frac{\alpha e^{-(\lambda t)^\beta}}{1 - \bar{\alpha} e^{-(\lambda t)^\beta}}, \quad t \in \mathcal{R}^+.$$

For fixed $t > 0$, let us define the function $l(\lambda) = \log \bar{F}(t; \lambda)$, $\lambda > 0$. It is then easy to observe that

$$l''(\lambda) = -(\beta - 1) \lambda^{\beta-2} - \frac{(\beta - 1) \bar{\alpha} \lambda^{\beta-2} \left(1 - \bar{\alpha} e^{-(\lambda t)^\beta}\right) - \beta \bar{\alpha} t^\beta \lambda^{2\beta-1} e^{-(\lambda t)^\beta}}{\left(1 - \bar{\alpha} e^{-(\lambda t)^\beta}\right)^2},$$

which is non-negative for any $0 < \alpha \leq 1$, $0 < \beta \leq 1$, and so $l(\lambda)$ is convex with respect to λ for any $0 < \alpha \leq 1$, $0 < \beta \leq 1$. Now, from Proposition A.6 of (Marshall and Olkin (2007), p. 690), we can conclude that for any $0 < \alpha \leq 1$, $0 < \beta \leq 1$ and fixed $t > 0$, $\bar{F}(t; \lambda)$ is convex with respect to λ . Moreover, we can readily see that $\bar{F}(t; \lambda)$ is decreasing with respect to λ . These observations confirm Condition (i) of Theorem 4.2. In addition, from Corollary 3.4, we readily observe that the survival function of $\sum_{i=1}^n X_{\lambda_i}$ is Schur-convex in $\boldsymbol{\lambda}$ for any $0 < \alpha \leq 1$. So, Condition (ii) of Theorem 4.2 is also satisfied, which completes the proof of the theorem. \square

References

- Barmalzan, G., Payandeh Najafabadi, A.T. and Balakrishnan, N. (2015). Stochastic comparison of aggregate claim amounts between two heterogenous portfolios and its applications, *Insurance: Mathematics and Economics*, **61**, 235-241.
- Coriderio, G.M. and Lemonte, A.J. (2013). On the Marshall-Olkin extended Weibull distribution, *Journal of Statistical Papers*, **54**, 333-353.
- Denuit, M. and Frostig, E. (2006). Heterogeneity and the need for capital in the individual model, *Scandinavian Actuarial Journal*, **1**, 42-66.
- Frostig, E. (2001). A comparison between homogeneous and heterogeneous portfolios, *Insurance: Mathematics and Economics*, **29**, 59-71.
- Ghitany, M.E., Al-Hussaini, E.K. and AlJarallah, R.A. (2005). Marshall-Olkin extended Weibull distribution and its application to censored data, *Journal of Applied Statistics*, **32**, 1025-1034.

- Hu, T. (1995). Monotone coupling and stochastic ordering of order statistics, *System Science and Mathematical Sciences*, **8**, 209-214.
- Hu, T. and Ruan, L. (2004). A note on multivariate stochastic comparisons of Bernoulli random variables, *Journal of Statistical Planning and Inference*, **126**, 281-288.
- Kaas, R. and Hesselager, O. (1995). Ordering claim size distribution and mixed Poisson probabilities, *Insurance: Mathematics and Economics*, **17**, 193-201.
- Kaas, R., Van Heerwaarden, A.E. and Goovaerts, M.J. (1994). Ordering of actuarial risks. *In: Caire Education Series 1*, Brussels, Belgium.
- Karlin, S. and Novikoff, A. (1963). Generalized convex inequalities, *Pacific Journal of Mathematics*, **13**, 1251-1279.
- Khaledi, B. and Ahmadi, S.S. (2008). On stochastic comparison between aggregate claim amounts, *Journal of Statistical Planning and Inference*, **138**, 3121-3129.
- Ma, C. (2000). Convex orders for linear combinations of random variables, *Journal of Statistical Planning and Inference*, **84**, 11-25.
- Marshall, A.W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika*, **84**, 641-652.
- Marshall, A.W. and Olkin, I. (2007). *Life Distributions*, Springer-Verlag, New York.
- Marshall, A.W., Olkin, I. and Arnold, B.C. (2011). *Inequalities: Theory of Majorization and its Applications*, Second edition, Springer, New York.
- Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*, Springer, New York.



On the analysis of the two parameter exponential distribution based on progressive type II censored data

R. Arabi Belaghi, H. Bevrani, M. Mohammadi¹

Department of Statistics, Faculty of Mathematical Sciences, University of Tabriz,
Tabriz, Iran

Abstract: In this study, the estimation of the two parameters exponential distribution based on progressive type II censored data were considered. The maximum likelihood (ML), penalized maximum likelihood (PML) and Bayes estimators assuming both of the location and scale parameters to be unknown were obtained. The results show that the PMLE is the same as uniformly minimum variance unbiased estimator (UMVUE). The mean square errors of proposed estimators both analytical and a Mont Carlo simulation study for different types of censoring schemes were computed. The simulation results revealed that the Bayes estimators outperforms the PMLEs. Further the PMLEs is superior to the MLEs.

Keywords: Bayes Estimators, Maximum Likelihood Estimators, Penalized Maximum Likelihood Estimators, Progressive Type II Censored Data.

Mathematics Subject Classification (2010): 62N01 62N02.

1 Introduction

The two parameters exponential distribution has many real world applications. It can be used to model the data such as service time of agents in a system (Queuing Theory), the time it takes before your next telephone call, the time until a radioactive particle decays, the distance between mutations on a DNA strand, and the extreme values of annual snowfall or rainfall. The probability density function

¹M. Mohammadi : mojtahamohamadi8@gmail.com

(p.d.f) of the two parameters exponential distribution is given by

$$f_{\eta,\theta}(x) = \frac{1}{\theta} e^{-\frac{(x-\eta)}{\theta}}, \quad x > \eta, \theta > 0, \eta \in R. \quad (1.1)$$

Due to its importance the estimation of the parameters of this distribution have been considered by many authors. [Lawless \(1982\)](#) and [Mann et al. \(1974\)](#) considered the earlier studied upon the estimation of exponential parameters. More recent works can also be found in [Johnson et al. \(1982\)](#).

In reliability and survival analysis studies, the experimenter may not always obtain complete information on failure times for all experimental units. Therefore, the censored data obtained from such experiments and censoring occurs commonly. Type-I and type-II censoring schemes are the two most common and popular censoring schemes. Briefly, they can be described as follows; consider n items under observation in a particular experiment. In type-I censoring scheme, the experiment continues up to a pre-specified time T and the number of observed failure is a random variable. On the other hand, in type-II censoring scheme requires the experiment to continue until a pre-specified number of observed failures occur, and the experimental time is a random variable. Recently progressive censoring scheme introduced has gained special attention in theoretical and applied statistics due to the flexibility in removing the pre-specified number of item during the test. Suppose n items are put in the life test. At the time of the first failure, R_1 of the surviving $n - 1$ items are removed randomly from the test. Then after the second failure R_2 out of $n - R_1 - 2$ remaining survived items are removed. This test continued until the m^{th} failure occur. At this stage, all the remaining $R_m = n - R_1 - R_2 - \dots - R_{m-1} - m$ survived units withdrawn from the test. Which m and $\mathbf{R} = (R_1, R_2, \dots, R_m)$ are prefixed. Many inferences and studies have been done based on progressive censored data. One may refer to [Cohen \(1991\)](#), [Aggarwala and Balakrishnan \(1998\)](#), [Balakrishnan and Cramer \(2014\)](#) and [Gurunlu Alma and Arabi Belaghi \(2015c\)](#) for more details.

In section 2 the maximum likelihood, Penalized maximum likelihood and Bayes estimators for both of the unknown parameters were considered. Section 3 contains the simulation study for assessing the performance of the proposed estimators. At the end of the paper a summary and conclusion are presented.

2 Proposed Estimators

In this section we obtain the Maximum likelihood (ML) and penalized maximum likelihood PML estimators of the parameters based on Progressive Type II censored data. Let $x_1 = x_{1:m:n}, x_2 = x_{2:m:n}, \dots, x_m = x_{m:m:n}$ are the progressive Type II censored data from a continuous population X with the p.d.f $f(\cdot)$ and c.d.f $F(\cdot)$, then the likelihood of the observations is given as

$$L(\theta, \eta) = A \prod_{i=1}^m f(x_i, m, n) [1 - F(x_i, m, n)]^{r_i}, \quad (2.1)$$

where

$$A = n(n - r_1 - 1) \cdots (n - \sum_{i=1}^{m-1} (1 + r_i)).$$

2.1 ML Estimators

For obtaining the MLE,

$$L(\theta, \eta) = A \frac{1}{\theta^m} e^{-\frac{1}{\theta} \sum_{i=1}^m (1+r_i)(x_i - \eta)} \quad ; x_i \geq \eta. \quad (2.2)$$

And logarithm of the likelihood function is:

$$\ln L(\theta, \eta) = \ln A - m \ln \theta - \frac{1}{\theta} \sum_{i=1}^m (1 + r_i)(x_i - \eta); \quad x_i \geq \eta \quad (2.3)$$

The log likelihood function is maximized with respect to η by taking $\eta = x_{(1)}$. To get the MLE for θ , the following equation be solved.

$$\frac{\partial}{\partial \theta} \ln L(\theta, \eta) = -\frac{m}{\theta^2} - \frac{1}{\theta^2} \sum_{i=1}^m (1 + r_i)(x_i - \hat{\eta}) = 0 \quad (2.4)$$

Therefore

$$\hat{\eta} = x_{(1)} \quad (2.5)$$

and

$$\hat{\theta} = \frac{1}{m} \sum_{i=1}^m (1 + r_i)(x_i - \hat{\eta}). \quad (2.6)$$

Theorem 2.1. Let $x_{1:m:n}, \dots, x_{m:m:n}$ be order statistics form $EXP(\theta, \eta)$. Then

$$Z_1 = n(x_{1:m:n} - \eta), \dots, Z_m = (n - \sum_{i=1}^{m-1} r_i - m)(x_{m:m:n} - x_{m-1:m:n}), \quad (2.7)$$

$1 \leq m \leq n - 1$ are independent and identically distributed with common distribution $EXP(\theta)$.

Proof. The proof can be found in [Aggarwala and Balakrishnan \(1998\)](#). \square

Now, by making use of [2.1](#), the expectation, bias and the variance of $\hat{\theta}$ and $\hat{\eta}$ are as follows

$$E(\hat{\theta}) = \theta - \frac{\theta}{m}, \quad (2.8)$$

$$Var(\hat{\theta}) = \frac{m-1}{m^2} \theta^2, MSE(\hat{\theta}) = \frac{\theta^2}{m}. \quad (2.9)$$

2.2 Penalized MLE

The most commonly used method of estimation is Maximum Likelihood Estimation (MLE). Although, under some regularity conditions, the MLE method has proper properties such as consistency and efficiency. But is too conservative because it always chooses the minimum of the sample to estimate the location parameter. The penalized maximum likelihood estimators for both parameters will be UMVUE.

$$L^*(\theta, \eta) = A(x_1 - \eta) \frac{1}{\theta^m} e^{-\frac{1}{\theta} \sum_{i=1}^m (1+r_i)(x_i - \eta)} \quad ; x_i \geq \eta \quad (2.10)$$

Logarithm of the likelihood function is

$$\ln L^*(\theta, \eta) = \ln A + \ln(x_1 - \eta) - m \ln \theta - \frac{1}{\theta} \sum_{i=1}^m (1+r_i)(x_i - \eta); x_i \geq \eta \quad (2.11)$$

By solving the following equations

$$\begin{cases} \frac{\partial}{\partial \theta} \ln L^*(\theta, \eta) = -\frac{m}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^m (1+r_i)(x_i - \eta) = 0 \\ \frac{\partial}{\partial \eta} \ln L^*(\theta, \eta) = -\frac{1}{x_{(1)} - \eta} + \frac{1}{\theta} \sum_{i=1}^m (1+r_i) = 0, \end{cases} \quad (2.12)$$

It is easy to verify that the PMLEs are:

$$\Rightarrow \begin{cases} \theta^* = \frac{1}{m-1} \sum_{i=1}^m (1+r_i)(x_{(i)} - x_{(1)}) \\ \eta^* = x_{(1)} - \frac{\theta^*}{n} \end{cases} \quad (2.13)$$

Considering the PMLEs are exactly the UMVUE! Now, by using 2.1, the expectation of θ^* is

$$E(\theta^*) = \frac{1}{m-1} E\left(\sum_{i=1}^m (1+r_i)(x_{(i)} - x_{(1)})\right) = \frac{1}{m-1} (m-1)\theta = \theta.$$

Thus θ^* is unbiased estimator of θ . Also, the variance of θ^* is:

$$Var(\theta^*) = \frac{1}{(m-1)^2} (m-1)\theta^2 = \frac{\theta^2}{m-1}.$$

Consider that $x_{(1)}$ and θ^* are independent, and the expectation of η^* is given by

$$E(\eta^*) = E(x_{(1)}) - \frac{E(\theta^*)}{n} = \frac{\theta}{n} + \eta - \frac{\theta}{n} = \eta.$$

So η^* is an unbiased estimator of η . The η^* variance is:

$$Var(\eta^*) = \frac{m\theta^2}{n^2(m-1)}.$$

2.3 Bayes Estimators

In this part the estimation in the Bayes context will be considered. In this regard, the following prior distribution to the parameters will be applied as follows:

$$\pi_1(\eta|\theta) \propto \frac{1}{\theta}$$

and

$$\pi_2(\theta) \propto \theta^{-(\alpha+1)} e^{-\frac{\beta}{\theta}}, \quad \theta > 0, \alpha > 0, \beta > 0$$

The posterior distribution can easily be found as:

$$\pi(\theta, \eta|x) \propto \frac{1}{\theta^{m+\alpha+2}} e^{-\frac{1}{\theta}[\sum_{i=1}^m (1+r_i)(x_i - \eta) + \beta]}, \quad \theta > 0, \eta \leq x_{(1)}.$$

By taking the marginal integration, the marginal posterior distribution can be computed as:

$$\pi(\theta|x) = \int \pi(\theta, \eta) d\eta$$

Thus,

$$\pi(\theta|x) \propto \frac{1}{\theta^{m+\alpha+1}} e^{-\frac{1}{\theta}[\sum_{i=1}^m (1+r_i)(x_i - x_{(1)}) + \beta]}.$$

It is easy to verify the marginal posterior distribution of θ given x is an Inverse gamma. That is:

$$\theta|x \sim IG(m + \alpha, \sum_{i=1}^m (1 + r_i)(x_i - x_{(1)}) + \beta).$$

Thus, under the squared error loss (SEL), the Bayes estimators of the scale parameter can be found as:

$$\delta_{\theta}^{SLE} = E(\theta|x) = \frac{\sum_{i=1}^m (1 + r_i)(x_i - x_{(1)}) + \beta}{m + \alpha - 1},$$

With the similar discussion, the posterior distribution of η given x can be computed as:

$$\pi(\eta|x) = \frac{n(m + \alpha) \left(\sum_{i=1}^m (1 + r_i)(x_i - x_{(1)}) + \beta \right)^{m+\alpha}}{\left(\sum_{i=1}^m (1 + r_i)(x_i - \eta) + \beta \right)^{m+\alpha+1}},$$

Therefore under SEL the Bayes estimator of the location parameter can be obtained as

$$\delta_{\eta}^{SLE} = E(\eta|x) = x_{(1)} - \frac{\sum_{i=1}^m (1 + r_i)(x_i - x_{(1)}) + \beta}{n(m + \alpha - 1)}.$$

It is obvious that the risk of the proposed Bayes estimators can not be achieved analytically. Consequently, a simulation study will be considered in next section.

3 Simulation Study

In order to assess the performance of proposed estimators a Monte Carlo simulation study for different values of n , m and R will be considered. It is obvious that the MSEs of the MLE and PMLE are free from \mathbf{R} which is ambiguous in practical situation to choose appropriate progressive censoring scheme. The Bayes estimators of the parameters can not be computed analytically, so the risk of the estimators is computed by simulation through following steps:

Algorithm 1. • Based on the Algorithm in [Balakrishnan and Sandhu \(1995\)](#), Generate the progressive Type II censored sample for the given values of R , n , m , η and θ .

- Compute The values of the proposed estimators based on the sample in step 1.

- Repeat the previous steps for 1000 times and then obtain the average and variance of the estimators from 1000 repetitions.

3.1 Simulation results

The simulation results are given in table 1 based on the schemes while assuming $\alpha = \beta = 1$. For the $\hat{\eta}$ and $\hat{\theta}$ respectively, can be concluded.

- Generally, both the MSE and biases of Bayes estimators are less than MSE's and biases of MLEs. While the PMLEs has the least MSE and biases. Also the MSE's of various schemes are approximately the same. Further the Bayes estimators have always positive biases. Finally, as m/n increases, the values of MSEs decreases.
- The Bayes estimators have less biases than the MLEs but slightly less than the PMLEs. Moreover for a fixed values of n , as m increases, the biases and MSEs become smaller. Finally, for a large sample size of n all proposed estimators have similar behaviors.

Conclusion

This paper considered the estimation of the parameters of two parameters exponential distribution based on Type II progressive censored data. The MLE, PMLE and Bayes estimators is especially proposed for both location and scale parameters under SEL function. The results indicated that the PMLE is the same as uniformly minimum variance unbiased estimator. The simulation study, also revealed that the PMLE and Bayes estimators are better than the MLEs.

References

- Aggarwala and Balakrishnan, N. (1998). *Some Properties of Progressive Censored Order Statistics From Arbitrary and Uniform Distributions with Applications to Inference And Simulation*, *Journal of Statistical Planning and Inference* **70**, 35- 49,

- Balakrishnan, N. and Sandhu, R., A. (1995). A Simple Simulation Algorithm for Generating Progressive Type-II Censored Samples, *American Statistician*, **49** (2), 229- 230.
- Balakrishnan, N. and Cramer, E. (2014). *Art of Progressive Censoring: Applications to Reliability and Quality*. Springer
- Cohen A. C. (1991). *Truncated and Censored Samples: Theory and Applications*. Marcel Dekker, New York.
- Gurunlu Alma, Ozlem, Arabi Belaghi, R. (2015c). On the estimation of Normal and Extreme Value distribution Parameters based on progressive Type II hybrid censored data, *Journal of Statistical Computation and Simulation*, Accepted. <http://dx.doi.org/10.1080/00949655.2015.1025785>.
- Mann N., R., Schafer, R., E., Singpurwalla, N., D., (1974). *Methods for Statistical Analysis of Reliability and Life Data*, New York, John Wiley and Sons.
- Johnson, N. L., Kots, S. , Balakrishnan, N. (1982), *Univariate Distribution* , vol 1, *Wiley*, New York.
- Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons, New York.

Table 1: progressive censor with n=50

$\theta = 0, \eta = 1$		<i>MLE</i>		<i>PMLE</i>		<i>Bayes</i>		
<i>m</i>	<i>r vector</i>	$\hat{\eta}$	$\hat{\theta}$	$\hat{\eta}$	$\hat{\theta}$	$\hat{\eta}$	$\hat{\theta}$	
8	$(42, 0^{*7})$	<i>Bias</i>	0.021	-0.1259	0.001	-0.0011	0.0009	-0.0009
		<i>MSE</i>	0.0009	0.128	0.0004	0.1465	0.0004	0.1121
	$(0^{*7}, 42)$	<i>Bias</i>	0.0195	-0.1224	-0.0005	-0.0029	-0.0005	0.0025
		<i>MSE</i>	0.008	0.1184	0.0004	0.135	0.0004	0.1033
	$(6, 5^{*6}, 6)$	<i>Bias</i>	0.0202	-0.1382	0.0002	0.0151	0.0004	-0.0132
		<i>MSE</i>	0.0008	0.1147	0.0004	0.1251	0.0004	0.0958
10	$(40, 0^{*9})$	<i>Bias</i>	0.0206	-0.0929	0.0006	0.0079	0.0005	0.0071
		<i>MSE</i>	0.0008	0.1038	0.0004	0.1175	0.0004	0.0952
	$(0^{*9}, 40)$	<i>Bias</i>	0.0195	-0.1016	-0.0005	-0.0017	-0.0004	-0.0015
		<i>MSE</i>	0.0008	0.0972	0.0004	0.1072	0.0004	0.0868
	(4^{*10})	<i>Bias</i>	0.0202	-0.1057	0.0002	0.0063	0.0003	-0.0057
		<i>MSE</i>	0.0008	0.0952	0.0004	0.1037	0.0004	0.084
15	$(35, 0^{*14})$	<i>Bias</i>	0.0209	-0.0828	0.0009	-0.0173	0.0012	-0.0161
		<i>MSE</i>	0.0009	0.0656	0.0004	0.0677	0.0004	0.059
	$(0^{*14}, 35)$	<i>Bias</i>	0.0194	-0.0739	0.0005	-0.0078	-0.0004	-0.0072
		<i>MSE</i>	0.0007	0.0654	0.0003	0.0689	0.0003	0.06
	$(5, 2^{*13}, 4)$	<i>Bias</i>	0.0207	-0.06666	0.0007	$5.13e - 06$	0.0006	$4.7e - 06$
		<i>MSE</i>	0.0008	0.0708	0.0004	0.0762	0.0004	0.0664



A method for selecting generating OWA operator weights models

Alireza Chaji ³

Shohadaie Hovaizeh University of Technology, khuzestan, Iran

Abstract: In this paper, a new model based on a measure of entropy for obtaining the ordered weighted averaging (OWA) operators is introduced. In the model it is assumed that, according to the available information, the OWA weights satisfy inequality constraints. The model is solved according to some given constraints with specific level of orness comparing the results with other methods. The results demonstrate the efficiency of our model in generating the OWA operator.

Keywords: OWA operator, Operator weights, Maximum entropy.

Mathematics Subject Classification (2010): 94A17 47bxx 47B37.

1 Introduction

The new type of operators introduced by Yager (1988) for aggregation is called as ordered weighted aggregation (OWA) operator. Indeed, these new OWA operators are as a family of mean type aggregation operators. OWA operators have many applications in various fields such as: decision making, approximate reasoning, neural network, expert systems, fuzzy system and control, data mining and linguistic systems (Merigó, José. M and Casanovas 2010, Kacprzyk, Zadrozny 2001; Ogryczak, Liwski 2003; Peláez, Doña 2003; Liu 2006; Ribeiro, Pereira 2003; Torra 2004; Yager 1988, 2004, 2009; Herrera-Viedma et al. 2003, 2006, 2007; Porcel et al. 2009 and Herrera-Viedma and Pasi 2003). Yager (1988) defined the degrees of "orness" and "andness" of the OWA operators and the measure of the dispersion of OWA weights by the measure of entropy. O'Hagan (1988) proposed a maximum entropy method subject to a specific level of orness, naming the obtained weights as maximum entropy OWA

³ Alireza Chaji : chajialireza@iust.ac.ir

(MEOWA) operators. Fuller and Majlender (2000) used the method of Lagrange multipliers to obtain an analytical form of the maximum entropy OWA operators. Majlender (2005) introduced a maximum Rényi entropy to generate the OWA operators, this model contained MEOWA operators. Parametric maximum entropy OWA (PMEOWA) weights are determined by Liu and Chen (2004). A new measure of entropy was presented by Yager (1995). Wu et al. (2009) used this measure and proposed a maximal Yager's entropy model for obtaining the OWA operators.

A more general form of entropy was defined by Kapur (1992) and applies when we have additional information about our problem. A decision maker may have inequality constraints for the OWA weights according to its available information about them. In order to use the additional information, we introduce a novel model based on the measure of entropy to determine the OWA operator weights. We solve the model for some constraints with different levels of orness and compare the results with other associated models.

The rest of the paper is organized as follows: Section 2 gives the definition of the OWA operators and an overview of the main existing models for obtaining the OWA weights based on the entropy. Section 3 proposes the maximum entropy model for determining the OWA operator weights when they satisfy inequality constraints. In section 4, the new model is solved for a number of given information with specific levels of orness. In the last section the paper is concluded.

2 Obtaining the OWA weights based on a measure of entropy when there are inequality constraints on the weights

In this section, we introduce a new measure of dispersion of the OWA weights based on a measure of entropy which defined by Kapur (1992). Utilizing this measure of entropy and proposing a novel model for determining the OWA operator weights.

In applied problems, on the basis of experiences or according to intuition, a decision maker may have reasons to assume that the OWA weights in the $W = [w_1, w_2, \dots, w_n]$ have to be required to satisfy the inequality constraints as:

$$x_i \leq w_i \leq y_i, \quad w_i \geq 0 \quad \text{for } i = 1, 2, \dots, n. \quad (2.1)$$

The inequality (2.1) can be written as:

$$w_i - x_i \geq 0, \quad y_i - w_i \geq 0.$$

Now we define a new measure of entropy of the OWA weights according to the inequality (2.1).

Definition 2.1. Assume W is an OWA weighting vector with elements w_1, w_2, \dots, w_n such that the weights are satisfy the inequality (2.1) then a new measure of entropy of the W is defined as:

$$\begin{aligned} S(W) = & -\sum_{i=1}^n (w_i - x_i) \ln(w_i - x_i), \\ & -\sum_{i=1}^n (y_i - w_i) \ln(y_i - w_i) - \sum_{i=1}^n (y_i - x_i) \ln(y_i - x_i). \end{aligned} \quad (2.2)$$

The definition 2.1 is permutationally symmetric in the sence that it does not change when the triplets (w_i, x_i, y_i) are permuted among themselves.

Using this measure of entropy and the principle of maximum entropy, we introduce the following model for obtaining the OWA operator weights with a specific level of orness as:

$$\begin{aligned} S(W) = & -\sum_{i=1}^n (w_i - x_i) \ln(w_i - x_i), \\ & -\sum_{i=1}^n (y_i - w_i) \ln(y_i - w_i) - \sum_{i=1}^n (y_i - x_i) \ln(y_i - x_i), \\ \text{subject to :} & \\ \text{orness}(w) = & \sum_{i=1}^n \frac{n-i}{n-1} w_i = \alpha \quad 0 \leq \alpha \leq 1, \\ x_i \leq w_i \leq y_i, & \quad \sum_{i=1}^n w_i = 1, \quad w_i \geq 0, \quad i = (1, 2, \dots, n). \end{aligned} \quad (2.3)$$

We have the following theorems from the proposed model.

Theorem 2.2. The maximum value of the measure of entropy $S(x)$ occurs when $y_i - w_i = w_i - x_i$ for all i .

In the following, we use the Lagrange multiplier method to obtain an analytic solution for determining the MBEOWA operator weights. This will enables us to analysis the MBEOWA weights deeper and prove some properties of the weights as well as simplify the process used for obtaining the MBEOWA weights.

The Lagrange function of the objective function subject to the constraints of the model where λ, γ and $x_i = 0$ are real numbers is as following:

$$L = \sum_{i=1}^n w_i \ln w_i + \sum_{i=1}^n (y_i - w_i) \ln (y_i - w_i) - \sum_{i=1}^n y_i \ln y_i + \lambda_1 (\sum_{i=1}^n \frac{n-i}{n-1} w_i - \alpha) + \lambda_2 (\sum_{i=1}^n w_i - 1).$$

Taking the partial derivatives of L with respect to w_i, λ_1 and λ_2 and setting them equal to zero so we obtain:

$$\frac{\partial L}{\partial w_i} = \ln\left(\frac{y_i - w_i}{w_i}\right) + \lambda_1 \frac{n-i}{n-1} + \lambda_2 = 0, \quad \text{for } i = 1, 2, \dots, n \tag{2.4}$$

$$\frac{\partial L}{\partial \lambda_1} = \sum_{i=1}^n \frac{n-i}{n-1} w_i - \alpha = 0, \tag{2.5}$$

$$\frac{\partial L}{\partial \lambda_2} = \sum_{i=1}^n w_i - 1 = 0. \tag{2.6}$$

From (2.4) and (2.6) we have

$$w_i = \frac{y_i}{1 + \exp(-\lambda_1 \frac{n-i}{n-1} - \lambda_2)}, \quad i = 1, 2, \dots, n$$

$$\sum_{i=1}^n w_i = \sum_{i=1}^n \frac{y_i}{1 + \exp(-\lambda_1 \frac{n-i}{n-1} - \lambda_2)} = 1.$$

Thus,

$$w_i = \frac{\frac{y_i}{1 + \exp(-\lambda_1 \frac{n-i}{n-1} - \lambda_2)}}{\sum_{i=1}^n \frac{y_i}{1 + \exp(-\lambda_1 \frac{n-i}{n-1} - \lambda_2)}}, \quad i = 1, 2, \dots, n. \tag{2.7}$$

and

$$\alpha_n(\lambda_1, \lambda_2) = \text{orness}(w) = \frac{1}{n-1} \sum_{i=1}^n (n-i) \frac{\frac{y_i}{1 + \exp(-\lambda_1 \frac{n-i}{n-1} - \lambda_2)}}{\sum_{i=1}^n \frac{y_i}{1 + \exp(-\lambda_1 \frac{n-i}{n-1} - \lambda_2)}}. \tag{2.8}$$

For $i = n$ and $i = 1$ from equation (2.4) we have

$$\ln \frac{y_n - w_n}{w_n} + \lambda_2 = 0 \iff \lambda_2 = \ln \frac{w_n}{y_n - w_n},$$

$$\ln \frac{y_1 - w_1}{w_1} + \lambda_1 + \lambda_2 = 0 \iff \lambda_1 = \ln \frac{w_1}{y_1 - w_1} - \lambda_2.$$

Thus,

$$\lambda_1 = \ln \frac{w_1 (y_n - w_n)}{w_n (y_1 - w_1)}.$$

Therefore, from (1-4) for any i we can find

$$w_i = \frac{y_i}{1 + \exp\left(\frac{n-i}{n-1} \ln \frac{w_n(y_1-w_1)}{w_1(y_n-w_n)} - \ln \frac{w_n}{y_n-w_n}\right)}$$

and

$$\ln\left(\frac{w_i}{y_i - w_{i_i}}\right) = \ln \frac{w_1(y_n - w_n)}{w_n(y_1 - w_1)} \frac{n-i}{n-1} + \ln \frac{w_n}{y_n - w_n}.$$

Thus,

$$\ln \frac{w_i}{\beta_i} = \frac{i-1}{n-1} \ln \frac{w_n}{\beta_n} + \frac{n-i}{n-1} \ln \frac{w_1}{\beta_1}$$

and

$$w_i = \beta_i^{(n-1)} \sqrt{\left(\frac{w_n}{\beta_n}\right)^{i-1} \left(\frac{w_1}{\beta_1}\right)^{n-i}}. \quad (2.9)$$

The minimum value of the $S(x)$ is equal to zero when each $w_i = x_i$ or $w_i = y_i$.

Theorem 2.3. *Suppose that the OWA weights obtained from the model (2.3) with the orness(W^*) = α when $a_1 \leq w_1 \leq b_1, \dots, a_n \leq w_n \leq b_n$ are $W^* = (w_1^*, w_2^*, \dots, w_n^*)$. Then, $\hat{W}^* = (\hat{w}_1^*, \hat{w}_2^*, \dots, \hat{w}_n^*)$ is the optimal solution of the model (7) for the orness(\hat{W}^*) = $1 - \alpha$ such that $\hat{w}_i^* = w_{n-i+1}^*$ when $a_{n-i+1} \leq \hat{w}_i \leq b_{n-i+1}$.*

For an example of Theorem 2.3, suppose that in the constrain problem (2.3) $0.1 \leq w_1 \leq 0.25, 0.1 \leq w_2 \leq 0.25, 0.1 \leq w_3 \leq 0.25, 0.2 \leq w_4 \leq 0.35$ and $0.0 \leq w_5 \leq 0.35$. We can get the optimum solutions $W = (0.131, 0.148, 0.169, 0.291, 0.259)$ for the specific levels of orness, $\alpha = 0.4$. Also, when $0.1 \leq w_5 \leq 0.25, 0.1 \leq w_4 \leq 0.25, 0.1 \leq w_3 \leq 0.25, 0.2 \leq w_2 \leq 0.235, 0.0 \leq w_1 \leq 0.35$ we have $W = (0.259, 0.291, 0.166, 0.148, 0.131)$ for $\alpha = 0.6$.

3 Numerical illustrations

The introduced model can be used in problems which we have inequality constraints on the OWA weights. we provide numerical examples in order to demonstrate the efficiency of the introduced model for obtaining the OWA operator weights.

Table 1: The OWA operator weights generated by the proposed model and some related models with degree of orness(W)=0.9

W	The proposed method	χ^2 method (CSM)	Least squares method (LSM)	Maximum entropy method (MEN)	Minimum variance method (MVM)	Yager's entropy method
w_1	0.375	0.406	0.353	0.369	0.360	0.400
w_2	0.275	0.255	0.292	0.257	0.280	0.200
w_3	0.175	0.153	0.200	0.167	0.200	0.200
w_4	0.125	0.103	0.107	0.108	0.120	0.200
w_5	0.050	0.082	0.046	0.070	0.040	0.000
Entropy	0.483	-	-	1.435	-	0.640

Example 3.1. Suppose $n = 5$, $\alpha = 0.7$ and the given constraints on the OWA weights are $0.0 \leq w_1 \leq 0.45$, $0.15 \leq w_2 \leq 0.35$, $0.05 \leq w_3 \leq 0.35$, $0.05 \leq w_4 \leq 0.2$ and $0.01 \leq w_5 \leq 0.15$ and thus, the optimal solution of the constrain problem (2.3) are shown in the Table 1.

Example 3.2. Assume orness (W)=0.8 and $0.1 \leq w_1 \leq 0.25$, $0.1 \leq w_2 \leq 0.25$, $0.1 \leq w_3 \leq 0.25$, $0.2 \leq w_4 \leq 0.35$ and $0.0 \leq w_5 \leq 0.35$ the obtained OWA operators are illustrated in Table 2

The model is solved by using the LINGO software package. Furthermore, for comparison: the MEOWA weights, the maximal Yager's entropy weights, the χ^2 method (CSM), Least squares method (LSM) and the minimum variance method (MVM) are shown in the tables.

It is observed from tables 1 and 2 that if the length of the intervals for the OWA weights decreasing then the entropy of them is decreasing. Therefore, having information about the boundaries of the weights is important for determining them. Virtually, in this situation we have smaller uncertainty of the weights comparing with the previous models which supposed that the weights are only positive.

Table 2: The OWA operator weights generated by the proposed model and some related models with degree of orness(W)=0.9

W	The pro- posed method	χ^2 method (CSM)	Least squares method (LSM)	Maximum entropy method (MEN)	Minimum variance method (MVM)	Yager's entropy method
w_1	0.366	0.406	0.353	0.369	0.360	0.400
w_2	0.293	0.255	0.292	0.257	0.280	0.200
w_3	0.167	0.153	0.200	0.167	0.200	0.200
w_4	0.117	0.103	0.107	0.108	0.120	0.200
w_5	0.054	0.082	0.046	0.070	0.040	0.000
Entropy	0.707	-	-	1.435	-	0.640

Conclusions

In this paper, we proposed a novel model for determining the OWA operator weights based on a measure of entropy. Using available information of the corresponding intervals of the any OWA weight in the introduced model for obtaining the weights. Furthermore, some properties of the new model have been investigated. The effectiveness and applicability of the proposed model were illustrated by numerical examples.

References

Herrera-Viedma, E, Cordón. O, Luque. M, Lopez. A.G, Muñoz. A.M (2003). A model of fuzzy linguistic IRS based on multi-granular linguistic information, *International Journal of Approximate Reasoning*

- 34** 221-239.
- Herrera-Viedma. E, Cordón. O, Luque. M, Lopez. A.G, Muñoz. A.M (2007). A Model of Information Retrieval System with Unbalanced Fuzzy Linguistic Information, *International Journal Of Intelligent Systems* **22 11** 1197-1214.
- Herrera-Viedma. E, Pasi. G, López-Herrera. A.G, Porcel. C (2006). Evaluating the Information Quality of Web Sites: A Methodology Based on Fuzzy Computing with Words, *Journal of American Society for Information Science and Technology* **54 4** 538-549.
- Herrera-Viedma. E, Pasi. G (2003). Evaluating the Informative Quality of Documents in SGML-Format Using Fuzzy Linguistic Techniques Based on Computing with Words, *Information Processing and Management* **39 2** 233-249.
- Kacprzyk. J, Zadrozny. S (2001). Computing with words in intelligent database querying: standalone and internet-based applications, *Information Sciences* **134 1** 71-109.
- Kapur. J. N.(1992).Entropy Optimization Principles With Applications, *Academic Press*
- Liu. X(2006). Some properties of the weighted OWA operator, *IEEE Transactions on Systems, Man, and Cybernetics, Part B* **36 1** 118-127.
- Liu. X, Chen. H.(2004). On the properties of parametric geometric OWA operator, *International Journal of Approximate Reasoning* **35** 163-178.
- Merigó.José. M and Casanovas. Montserrat(2010). The Fuzzy Generalized OWA Operator and Its Application in Strategic Decision Making, *Cybernetics and Systems* **41 5** 359-370.
- Ogryczak. W, liwi ski. T. (2003). On solving linear programs with the ordered weighted averaging objective, *European Journal of Operational Research* **148** 80-91.
- O'Hagan. M (1988). Aggregating template rule antecedents in real-time expert systems with fuzzy set-logic, *In Proceedings of 22nd annual IEEE Asilomar conference on signals, systems, and computers Pacific Grove, CA* 681-689.

-
- Peláez. J.I, Doña. J.M (2003). Majority additive-ordered weighting averaging: a new neat ordered weighting averaging operator based on the majority process, *International Journal of Intelligent Systems* **18** 469-481.
- Ribeiro. R.A, Pereira. R.A.M (2003). Generalized mixture operators using weighting functions: a comparative study with WA and OWA, *European Journal of Operational Research* **142** 329-342.
- Torra. V (2004). OWA operators in data modeling and reidentification, *IEEE Trans. Fuzzy Systems* **12** 652-660.
- Yager. R. R (1995). Measures of entropy and fuzziness related to aggregation operators, *Information Sciences* **82** 147-166.
- Yager. R. R(2004). OWA aggregation over a continuous interval argument with applications to decision making, *IEEE Transactions on Systems, Man, and Cybernetics, Part B* **34** 1952-1963.



A Note on the Parrondo's Paradox

Ejlali, N¹, Pezeshk, H^{2,4} and Sadeghi, M^{3,4}.

¹Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, G.C., Tehran.

²School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran.

³National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran 14155-6346.

⁴School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran.

Abstract: Consider two losing games. When a player plays these games in a random or periodic strategy, one may expect to see a losing game. The Parrondo's paradox arises when this combined game leads to a winning one. The original game was invented by Parrondo as a discrete-time version of flashing Brownian ratchet. It has applied in several fields such as economics, physical quantum systems, and population genetics. In this paper, we introduce a new version of paradox by considering a new rule for one of those games. The paradoxical property is shown by some computer simulations.

Keywords Parrondo's paradox, losing game, winning game, randomized rule.

Mathematics Subject Classification (2010): 60J10 91A60.

1 Introduction

Parrondo's paradox was introduced by the Spanish physicist Juan M. R. Parrondo in 1996. It was presented in an unpublished form at a workshop in Torino, Italy. The paper was published by Harmer and Abbott in 1999. It has been applied in several fields such as economics [Spurgin \(2005\)](#) , physical

²Hamid Pezeshk : pezeshk@khayam.ut.ac.ir

quantum systems [Flitney \(2002\)](#), [Gawron \(2005\)](#) and population genetics [Reed \(2007\)](#).

The main idea of Parrondo's paradox is that two individually losing games, say, A and B, can be combined in a deterministic or non-deterministic mixing pattern to end up with a winning result. Different games have been introduced by several researchers. Parrondo et al. [Parrondo \(2000\)](#) introduced a new version for game B, where a player uses four different coins depending on his previous history of wins and losses. The effects of cooperation among players in Parrondos games have been considered by Toral [Toral \(2001\)](#). In the cooperative version of the paradox, the game A is the same as in the original paradoxical game, but in game B each of N players, arranged in a circle owns a capital $C_i(t)$, $i = 1, \dots, N$, will win the next play depending on the result of the previous run of the game for the neighbors. The average capital increases by combining two games [Toral \(2001\)](#). Zoran Mihailovic and Milan Rajkovic in 2006 extended the cooperative Parrondo's games on a two-dimensional lattice [Mihailovic \(2006\)](#).

Wayne Wah Ming Soo and Kang Hao Cheong in 2013 [Cheong \(2013\)](#) introduced a new approach to analyse the original paradox. They started with one process and tried to drive the complementary process.

In section two of this article, we review the original Parrondo's paradox. In section three, our new strategy for game B is introduced and some simulation results are presented. It is shown that combining two losing games by considering a threshold for the capital of the player in game B, will result in a fair game.

2 Original Parrondo's Games

The original Parrondo's games consist of two games, named, A and B. The game A is a simple coin tossing game where the capital of a player will increase (decrease) by one unit if heads (tails) occur. The probability of winning is denoted by p and the probability of losing is $q = 1 - p$. Game B consists of two sub-games with two coins and it depends on the capital. The probability of winning depends on the current capital of the player. If the capital is a multiple of M, the game will be played by the first coin and if it is not the game will be played by the second coin. Parrondo (1996) set the number

M equal to three and the probability of win is set to one of the p_1 or p_2 , therefore game B uses two different coins according to whether the capital of the player is a multiple of three or not.

For game A assume that f_j is the probability of the capital reaches zero in a finite number of plays, starting from a capital of j units. We have either

1) $f_j = 1$ for all $j \geq 0$, in this case the game is either fair or losing.

or

2) $f_j < 1$ for all $j > 0$, in this case there is some probability that the capital will increase so that the game is winning. (for more details see [Parrondo \(2000\)](#)).

The f_j can be obtained using the following recursive formula:

$$f_j = pf_{j+1} + (1-p)f_{j-1},$$

for $j \geq 0$ and the initial condition $f_0 = 1$.

The solution of the equation with this initial condition is

$$C_1 \left(\left(\frac{1-p}{p} \right)^j - 1 \right) + 1,$$

where C_1 is a constant, and the solution will be $\min\left(\left(\frac{1-p}{p}\right)^j, 1\right)$.

Therefore we can find that the game A is

$$\text{winning if } \frac{1-p}{p} < 1, \text{ i.e } p > 1/2,$$

$$\text{fair if } \frac{1-p}{p} = 1, \text{ i.e } p = 1/2,$$

$$\text{losing if } \frac{1-p}{p} > 1, \text{ i.e } p < 1/2.$$

As in the case of game A, for game B assume that g_j is the probability that the capital reaches zero in a finite number of plays, starting from a capital of j units. We have either

1) $g_j = 1$ for all $j \geq 0$, in this case the game is either fair or losing.

or

2) $g_j < 1$ for all $j > 0$, in this case there is some probability that the capital will increase so that the game is winning.

We can obtain g_j by the following recursive formulas:

$$\begin{aligned} g_{3j} &= p_1 g_{3j+1} + (1 - p_1) g_{3j-1}, \\ g_{3j+1} &= p_2 g_{3j+2} + (1 - p_2) g_{3j}, \\ g_{3j+2} &= p_2 g_{3j+3} + (1 - p_2) g_{3j+1}, \end{aligned}$$

for which the initial condition is $g_0 = 1$. So with this initial condition value the general solution will be:

$$C_2 \left(\left(\frac{(1 - p_1)(1 - p_2)^2}{p_1 p_2^2} \right)^j - 1 \right) + 1,$$

where C_2 is a constant. So the solution will be:

$$\min \left(\left(\frac{(1 - p_1)(1 - p_2)^2}{p_1 p_2^2} \right)^j, 1 \right),$$

Therefore we can find that the game B is

$$\begin{aligned} \text{winning if } & \frac{(1 - p_1)(1 - p_2)^2}{p_1 p_2^2} < 1, \\ \text{fair if } & \frac{(1 - p_1)(1 - p_2)^2}{p_1 p_2^2} = 1, \\ \text{losing if } & \frac{(1 - p_1)(1 - p_2)^2}{p_1 p_2^2} > 1. \end{aligned}$$

For more details see [Parrondo \(2000\)](#). Now we consider the combined game in which the game A is played with probability γ and the game B is played with probability $1 - \gamma$. If we show this combined game as $A + B$, the probabilities of winning and losing are as follows:

$$\text{Game } A + B = \begin{cases} p(\text{winning} | \text{the capital is multiple of three}) = \gamma p + (1 - \gamma) p_1, \\ p(\text{winning} | \text{the capital is not multiple of three}) = \gamma p + (1 - \gamma) p_2. \end{cases}$$

Infact the combined game has the form of the game B. The conditions are:

$$\begin{aligned} \text{winning if } & \frac{(1 - p_1^{(A+B)})(1 - p_2^{(A+B)})^2}{p_1^{(A+B)}(p_2^{(A+B)})^2} < 1, \\ \text{fair if } & \frac{(1 - p_1^{(A+B)})(1 - p_2^{(A+B)})^2}{p_1^{(A+B)}(p_2^{(A+B)})^2} = 1, \end{aligned}$$

$$\text{losing if } \frac{(1 - p_1^{(A+B)})(1 - p_2^{(A+B)})^2}{p_1^{(A+B)}(p_2^{(A+B)})^2} > 1,$$

where

$$p_1^{A+B} = \gamma p + (1 - \gamma)p_1,$$

$$p_2^{A+B} = \gamma p + (1 - \gamma)p_2.$$

The numerical values corresponding to the original Parrondo's games are $p = 1/2 - \varepsilon$, $p_1 = 1/10 - \varepsilon$ and $p_2 = 3/4 - \varepsilon$, where ε is a small noise parameter introduced to control the three probabilities. For a value of ε equal to zero, both games are fair, whereas if ε is small and positive both games are losing. In both cases, the combined game results in a winning game. see [Parrondo \(1996\)](#).

3 New Strategy and Simulation Results

In our model, game A is the same as the original game introduced by Parrondo, but we consider the new strategy for game B which depends on the capital of the last run. Our strategy for game B is defined as follows: we have two coins calling them as coin1 and coin2. We will compare current capital with some constant value, say C , and decide to choose one coin out of two coins to play the game.

Therefore if we denote the current capital in time t with $C(t)$, according to our strategy:

- 1) If $C(t) \geq C$, the probability of winning is assumed to be p_1 (i.e coin1 is tossed and the probability of heads is p_1).
- 2) If $C(t) < C$, the probability of winning is assumed to be p_2 (i.e coin2 is tossed and the probability of heads is p_2).

As mentioned above, game A is as the same in original game. It is a coin flipping or random walk as follows:

$$\text{GameA} = \begin{cases} p(\text{winning}) = \frac{1}{2} - \varepsilon. \\ p(\text{losing}) = \frac{1}{2} + \varepsilon. \end{cases}$$

Game B can be considered mathematically as follows:

$$\text{Game B} = \begin{cases} p(\text{winning}|C(t) \geq C) = p_1. \\ p(\text{losing}|C(t) \geq C) = 1 - p_1. \\ p(\text{winning}|C(t) < C) = p_2. \\ p(\text{losing}|C(t) < C) = 1 - p_2. \end{cases}$$

Now, we present some simulation results to show that the paradoxical property is present for different parameters in this case. For the case of game A the simulation is performed with probability $1/2 - \epsilon$. Apparently, the player will lose the game. In all simulations, games start with zero capital and the value for C is set to one. Different parameters are considered for game B and we combined two games periodically. Results are shown in the figure 1 and figure 2.

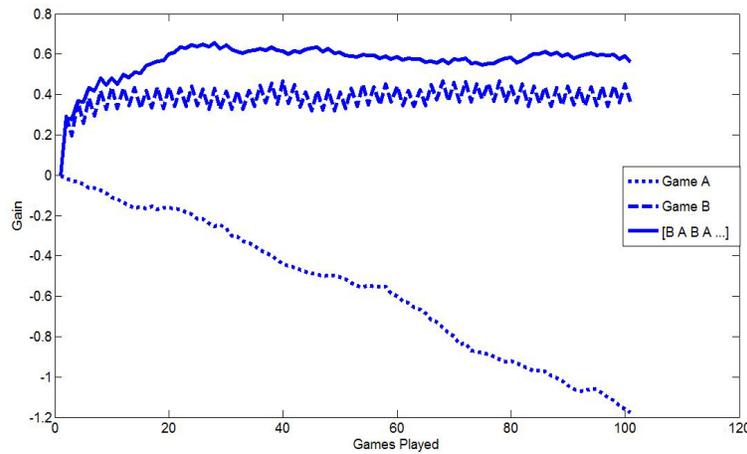


Figure 1: Figure of the gain versus the games played over 100 times. Game A is played with probability $p = 1/2 - \epsilon$ and game B is played with probabilities $p_1 = 0.35$ and $p_2 = 0.65$. Noise parameter is $\epsilon = 0.005$. Game A is a losing game and game B is a fair game. Two games are combined periodically, it is denoted by [B A B A ...]. Combined game is a fair game.

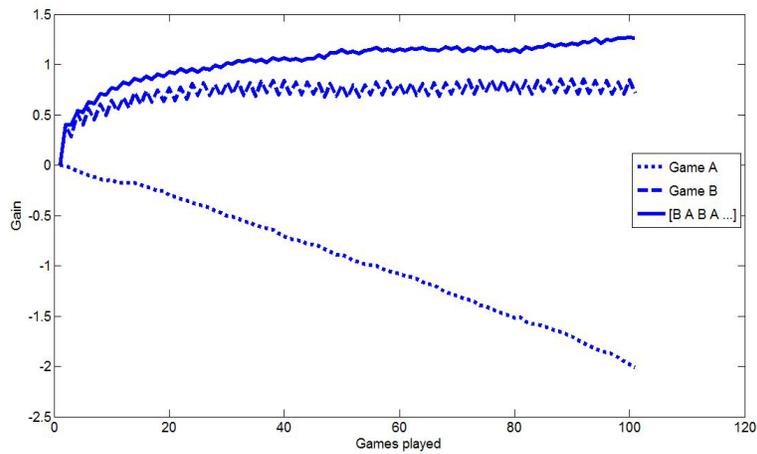


Figure 2: Figure of the gain versus the games played over 100 times. Game A is played with probability $p = 1/2 - \epsilon$ and game B is played with probabilities $p_1 = 0.35$ and $p_2 = 0.7$. Noise parameter is $\epsilon = 0.009$. Game A is a losing game and game B is a fair game. Two games are combined periodically, it is denoted by [B A B A ...]. Combined game is a fair game.

Conclusion

In this work, we introduce the new strategy for game B by considering the constant values as a threshold. We show that the paradox also occurs in this new model by switching between two games alternatively. Our simulation results show the occurrence of the paradox. Fair game is resulted by combining losing game A and fair game B.

Acknowledgements

Nasim Ejlali would like to thank Faculty of Mathematical Sciences of Shahid Beheshti University, Hamid Pezeshk and Mehdi Sadeghi were supported by a grant from school of biological sciences at IPM.

References

- Cheong, Kang Hao and Soo, Wayne Wah Ming.(2013). Construction of novel stochastic matrices for analysis of Parrondos paradox, *Physica A: Statistical Mechanics and its Applications*, **392(20)**, pp. 4727–4738.
- Flitney, Adrian P and Ng, Joseph and Abbott, Derek. (2002). Quantum Parrondo's games, *Physica A: Statistical Mechanics and its Applications*, **314(1)**, pp. 535–42.
- Gawron, Piotr and MISZCZAK, JAROSLAW A.(2005) Quantum implementation of Parrondo's paradox, *Fluctuation and Noise Letters*, **5(4)** pp. L471–L478.
- Harmer, Gregory P and Abbott, Derek and Taylor, Peter G and Parrondo, Juan MR. (2001). Brownian ratchets and Parrondos games, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **11(3)** , 705–714.
- Mihailović, Zoran and Rajković, Milan. (2006) Cooperative Parrondo's games on a two-dimensional lattice, *Physica A: Statistical Mechanics and its Applications*, **365(1)**, 244–251.
- Parrondo, JMR.(1996). How to cheat a bad mathematician, *EEC HC&M Network on Complexity and Chaos*.
- Parrondo, Juan MR and Harmer, Gregory P and Abbott, Derek. (2000) New paradoxical games based on Brownian ratchets, *Physical Review Letters*, **85(24)**, pp. 5226.
- Reed, Floyd A. (2007). Two-locus epistasis with sexually antagonistic selection: a genetic Parrondo's paradox, *Genetics*, **176 (3)**. pp. 1923–1929.
- Spurgin, Richard and Tamarkin, Maury. (2005). Switching investments can be a bad idea when Parrondo's paradox applies, *Computational Optimization and Applications*, **6(1)**, pp. 15–18.
- Toral, Raúl. (2001) Cooperative Parrondo's games, *Fluctuation and Noise Letters*, **1(01)**, pp. L7–L12.



A normal-Poisson distribution: Theory and applications

Fatemeh Esfandiyari^{1,2}, Eisa Mahmoudi²

^{1,2}Department of Statistics, Yazd University, Yazd, IRAN

Abstract: A new three-parameter distribution called the normal-Poisson (NP) distribution is introduced and studied in details. Some properties of the NP distribution is obtained such as moments and maximum likelihood estimators. The efficiency of this new distribution is illustrated through a real data set to show that it is quite flexible in comparing with the normal and skew normal distributions.

Keywords: Error function, Hazard function, Maximum likelihood estimation, Poisson distribution, Skew normal distribution.

Mathematics Subject Classification (2010): 60E05; 62H10; 62H12.

1 Introduction

In recent years, different techniques for extending the family of normal distributions have been proposed. The most usual procedure derived from [Azzalini \(1985\)](#) consists in deriving a skew-normal distribution through the transformation $f(x; \alpha) = 2\phi(x)\Phi(\alpha x)$, $x \in \mathbb{R}$ and $\alpha \in \mathbb{R}$, where $\phi(x)$ and $\Phi(x)$ denote the standard normal density and distribution, respectively. The method applied here can be considered an alternative to the well-known skew-normal distribution [Azzalini \(1985\)](#), whose properties ([Azzalini \(1986\)](#), [Azzalini and Chiogna \(2004\)](#)), estimation ([Gupta and Gupta \(2008\)](#)), diagnostics ([Xie et al. \(2009\)](#)), generalization ([Gupta and Gupta \(2004\)](#)) and multivariate extension ([Azzalini and Valle \(1996\)](#), [Azzalini and Capitanio \(1999\)](#)) have been widely developed.

Other ways of obtaining skewed normal distributions have also been introduced, such as the Balakrishnan skew-normal density in [Sharafi and Behboodian \(2008\)](#), the variance-gamma process in [Fung and Seneta \(2007\)](#) and the generalized normal distribution in [Nadarajah \(2005\)](#), among others.

²Speaker's Name: Fatemeh Esfandiyari, f.esfandiyari89@gmail.com

In this paper, we introduce a new generalization of the normal distribution $N(\mu, \sigma)$. This is done by adding a new parameter $\theta > 0$ to the normal distribution such that a skewness effect will depend on it. The method used to insert the new parameter to normal distribution is described in [Mahmoudi and Jafari \(2012\)](#), where it was applied to the generalized exponential distribution and introduced generalized exponential power series class of distributions. This method enables us to obtain explicit expressions for the probability density and survival and hazard rate functions. These expressions then allow us to ML-estimate the model parameters with the help of appropriate software.

To express the motivations of introducing this generalization of the normal distribution, following reasons might be referred to:

- As the failure rate function is studied, flexibility of distribution can be obviously observed in modeling different types of lifetime data. So it can be said that an important characteristic of the NP distribution is that its failure rate function can be increasing and unimodal-bathtub shaped.
- This distribution has considerable applications in other sciences and also routine issues such as calculating minimum waiting time to service customers in bank or store which in number of customers has poisson distribution and distribution of waiting time is normal.

2 The normal-Poisson Distribution

Suppose that the number of components in a system is a positive random variable where its pmf belongs to one of well-known discrete distributions. This problem can be seen in many natural phenomena such as queuing theory. Let there are N series components in a system and each component has the lifetime X_i , $i = 1, 2, \dots, N$, where the random variable N has Poisson distribution truncated at zero. The lifetime of such a system is denoted by $X = \min\{X_1, \dots, X_N\}$. We called this new generalization of the normal distribution as normal-Poisson (NP) distribution.

The cdf of NP distribution is given by

$$F(x; \mu, \sigma, \theta) = \frac{e^\theta - e^{\theta(1 - \Phi(\frac{x-\mu}{\sigma}))}}{e^\theta - 1}. \quad (2.1)$$

Differentiating from (2.1) w.r.t. x gives the pdf of NP distribution as follow:

$$f(x; \mu, \sigma, \theta) = \frac{\theta \phi\left(\frac{x-\mu}{\sigma}\right) e^{\theta(1-\Phi\left(\frac{x-\mu}{\sigma}\right))}}{\sigma(e^\theta - 1)}, \tag{2.2}$$

where parameters $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and $\theta \in \mathbb{R}$ is the shape parameter, which characterize the skewness, kurtosis, and unimodality of the distribution. A random variable X follows the NP distribution with parameters μ , σ and θ is denoted by $X \sim NP(\mu, \sigma, \theta)$. Plots of the NP density function for selected parameter values are given in Figure 1. An important

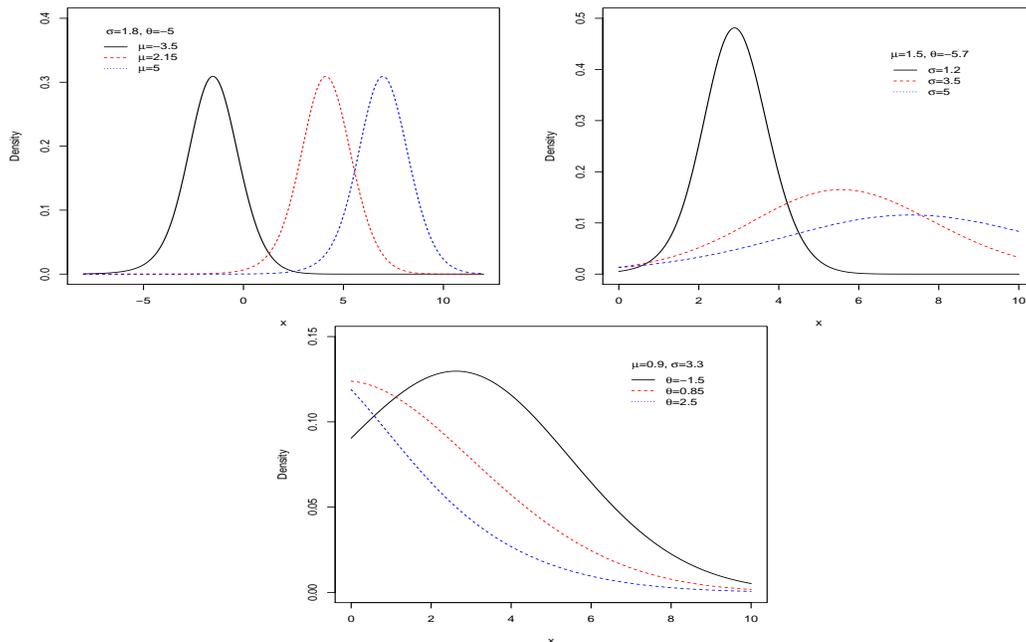


Figure 1: Plots of the NP density function for selected parameter values

characteristic of the NP distribution is that its density function can be unimodal that makes this distribution much advantages in modeling lifetime data.

2.1 Failure rate function and its shapes

The survival and failure rate functions of the NP distribution are given respectively by

$$S(y; \mu, \sigma, \theta) = \frac{e^{\theta(1-\Phi(\frac{y-\mu}{\sigma}))} - 1}{e^\theta - 1}, \quad (2.3)$$

and

$$h(y; \mu, \sigma, \theta) = \frac{\theta \phi(\frac{x-\mu}{\sigma}) e^{\theta(1-\Phi(\frac{x-\mu}{\sigma}))}}{\sigma(e^{\theta(1-\Phi(\frac{y-\mu}{\sigma}))} - 1)}. \quad (2.4)$$

Plots of the NP failure rate function for selected parameter values are given in Figure 2. An important

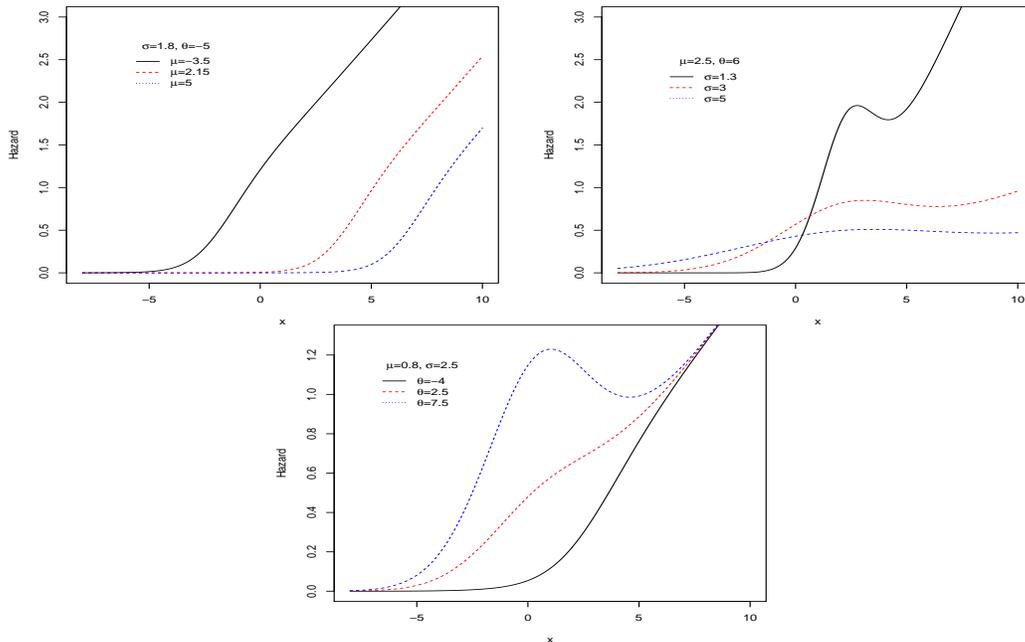


Figure 2: Plots of the NP failure rate function for selected parameter values

characteristic of the NP distribution is that its failure rate function can be increasing and unimodal-bathtub shaped that makes this distribution flexible in modeling different types of lifetime data.

2.2 Quantiles and Moments of the NP distribution

The p -th quantile of a NP distribution can be easily attained by the relationship

$$x_p = \mu + \sigma \Phi^{-1} \left[1 - \frac{\log(p(e^\theta - 1) + 1)}{\theta} \right].$$

We can use this expression for generating random data from NP with generating data from uniform distribution.

2.2.1 Moments

Let X be a random variable from $NP(\mu, \sigma, \theta)$, the k th moment can be obtained as

$$\mu_k = \int_0^1 (\mu + \sigma \Phi^{-1}(t))^k \frac{\theta e^{\theta(1-t)}}{e^\theta - 1} dt.$$

Thus, first three moments are as follows:

$$\begin{aligned} \mu_1 &= E(X) = \frac{\mu e^{-\theta}(1 + \sigma)}{e^\theta - 1} [1 - e^{-\theta}] \\ &\quad + \frac{\theta e^{-\theta}}{e^\theta - 1} \sqrt{2} \sigma^2 \times \int_0^1 \operatorname{erf}^{-1}(2t - 1) e^{-\theta t} dt, \\ \mu_2 &= E(X^2) = \frac{\theta e^{-\theta}}{e^\theta - 1} \left[\frac{\mu^2(1 + \sigma^2) + 2\mu\sigma}{\theta} [-e^{-\theta}] + 2\sqrt{2}\mu(\sigma^2 + \sigma^3) \right. \\ &\quad \left. \int_0^1 \operatorname{erf}^{-1}(2t - 1) e^{-\theta t} dt + 2\sigma^4 \int_0^1 [\operatorname{erf}^{-1}(2t - 1)]^2 e^{-\theta t} dt \right], \\ \mu_3 &= E(X^3) = \frac{\mu^3}{\theta} (1 - e^{-\theta}) [1 + \sigma^3 + 6\sigma] \\ &\quad + \int_0^1 \operatorname{erf}^{-1}(2t - 1) e^{-\theta t} dt \\ &\quad \left[3\sqrt{2}\mu^2\sigma^4 + 6\sigma^5\mu + 9\sqrt{2}\mu^4\sigma^3 + 6\sqrt{2}\mu^2\sigma^2 \right] \\ &\quad + 6\mu\sigma^3 \int_0^1 [\operatorname{erf}^{-1}(2t - 1)]^2 e^{-\theta t} dt + 2\sqrt{2}\sigma \int_0^1 [\operatorname{erf}^{-1}(2t - 1)]^3 e^{-\theta t} dt. \end{aligned}$$

Hence, variance and skewness coefficient can be readily derived from these moments. Table 1 contains values for μ_k , ($k = 1, 2, 3$) and the variance and the skewness coefficient in a $NP(\mu, \sigma, \theta)$ distribution for various values parameters.

Table 1: Moments, variance and the skewness coefficient for NP distribution

μ	σ	θ	μ_1	μ_2	μ_3	var	$skewness$
0	1	0.3	-0.0845	1.0041	-0.2113	0.9970	-0.2123
		0.6	-0.1684	1.0164	-0.4221	0.9881	-0.4297
		0.9	-0.2509	1.0366	-0.6311	0.9736	-0.6569
0.5	0.1	0.3	0.4915	0.2516	0.1335	0.0100	134.1181
		0.6	0.4832	0.2433	0.1272	0.0099	129.5082
		0.9	0.4749	0.2353	0.1211	0.0097	126.0544
-1	0.8	0.3	-1.0676	1.7779	-3.2390	0.6381	-6.3549
		0.6	1.1347	1.9199	-3.5717	0.6324	-7.1028
		0.9	-1.2007	2.0649	-3.9156	0.6231	-7.9607

3 Estimation and inference

Let X_1, \dots, X_n be observation from a random sample of size n from $NP(\mu, \sigma, \theta)$ and $\Theta = (\mu, \sigma, \theta)^T$ be the parameter vector. The log-likelihood function is given by

$$l_n(\Theta) = n \log(\theta) - n \log(\sigma(e^\theta - 1)) + \sum_{i=1}^n \log(\phi(\frac{x_i - \mu}{\sigma})) + \theta \sum_{i=1}^n (1 - \Phi(\frac{x_i - \mu}{\sigma})). \quad (3.1)$$

By differentiating with respect to μ, σ and θ , respectively, and setting the results equal to zero, the derivatives of (3.1) are given by

$$\frac{\partial l_n(\Theta)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) + \frac{\theta}{\sigma} \sum_{i=1}^n \phi(\frac{x_i - \mu}{\sigma}), \quad (3.2)$$

$$\frac{\partial l_n(\Theta)}{\partial \sigma} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 + \frac{\theta}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \phi(\frac{x_i - \mu}{\sigma}), \quad (3.3)$$

$$\frac{\partial l_n(\Theta)}{\partial \theta} = \frac{n}{\theta} - n(1 - e^{-\theta})^{-1} + \sum_{i=1}^n (1 - \Phi(\frac{x_i - \mu}{\sigma})). \quad (3.4)$$

The log-likelihood can be maximized either directly or by solving the nonlinear likelihood equations. One can apply Eqs. (3.2)-(3.4) in Matlab or R softwares to obtain the MLE of the parameters of NP distribution.

4 Applications

In this section, the NP distribution is fitted to a real data set and also compared the fitted NP with two relative models, Normal (N) and Skew-Normal (SN) distributions with pdf $\frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha\frac{x-\mu}{\sigma}\right)$, to show the superiority of the NP distribution.

The data set is given by [Birnbaum and Saunders \(1969\)](#) that refers to fatigue life of 6061-T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. The data set consists of 100 observations.

The TTT plot of this set of data in [Figure 3](#) displays an increasing hazard rate function. The MLEs of the parameters, $-2\log$ -likelihood, AIC (Akaike Information Criterion), the Kolmogorov-Smirnov test statistic (K-S), P-value are displayed in [Table 2](#) for this data set. The fitted densities and the empirical distribution versus the fitted cumulative distribution functions of NP, N and SN models are displayed in [Figure 4](#).

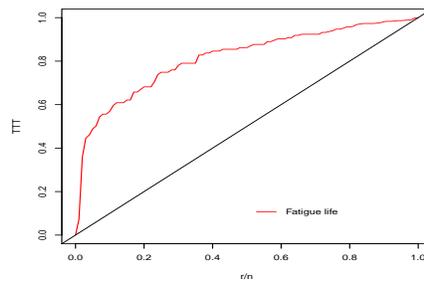


Figure 3: TTT plot of Birnbaum and Saunders data

The results for these data set show that the NP distribution yields the best fit among the N and SN distributions. For these data, the K-S test statistic takes the smallest value with the largest value of its respective p-value for NP distribution. Also, the values of the AIC for these distributions in [Table 2](#) and the plots of the densities and cumulative distribution functions in [Figure 4](#) confirmed this conclusion.

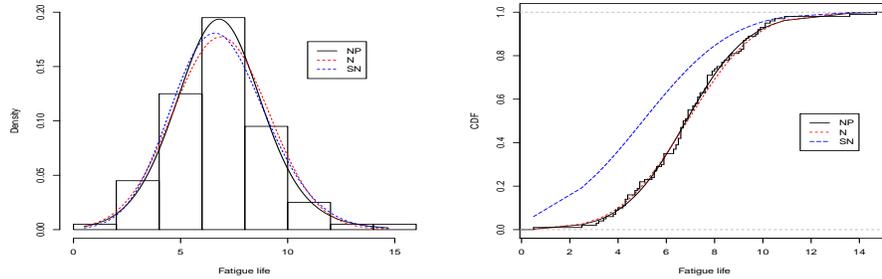


Figure 4: Plots of fitted NP, N and SN for fatigue life data.

Table 2: MLEs, -2 Log L, K-S, P-value and AIC for time between failures data.

Dist	MLE	-2 Log L	K-S	P-value	AIC
NP	$\hat{\mu} = 9.7762, \hat{\sigma} = 2.8162,$ $\hat{\theta} = 4.6477$	442.4137	0.05	0.9641	448.4137
N	$\hat{\mu} = 6.8780, \hat{\sigma} = 2.2499$	445.9741	0.0674	0.7535	449.9741
SN	$\hat{\mu} = 0.3993, \hat{\sigma} = 21.8577,$ $\hat{\alpha} = 1.3284$	65.0575	0.1204	0.8539	449.9741

References

- Azzalini, A., 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.
- Azzalini, A., 1986. Further results on a class of distributions which includes the normal ones. *Statistica*, **XLVI (2)**, 199-208.
- Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, Series B*, **61(3)**, 579-602.
- Azzalini, A., Chiogna, M., 2004. Some results on the stress-strength model for skew-normal variates. *Metron*, **3**, 315-326.
- Azzalini, A., Valle, A.D., 1996. The multivariate skew-normal distribution. *Biometrika*, **83(4)**, 715-726.

-
- Birnbaum, Z.W., Saunders, S.C., 1969. Estimation for a family of life distributions with applications to fatigue, *Journal of Applied Probability*, **6**, 328-347.
- Fung, T., Seneta, E., 2007. Tailweight, quantiles and kurtosis: a study of competing distributions. *Operations Research Letters*, **35**, 448-454.
- Gupta, R.C., Gupta, R.D., 2004. Generalized skew normal model. *Test*, **12(2)**, 501-524.
- Gupta, R.C., Gupta, R.D., 2008. Analyzing skewed data by power normal model. *Test*, **17(1)**, 197-210.
- Mahmoudi, E., Jafari, A.A., 2012. Generalized exponential-power series distributions. *Computational Statistics and Data Analysis*, **56**, 4047-4066.
- Nadarajah, S., 2005. A generalized normal distribution. *Journal of Applied Statistics*, **32(7)**, 685-694.
- Sharafi, M., Behboodian, J., 2008. The Balakrishnan skew-normal density. *Statistical Papers*, **49**, 769-778.
- Xie, F.G., Lin, J.G., Wei, B.C., 2009. Diagnostics for skew-normal nonlinear regression models with AR(1) errors. *Computational Statistics and Data Analysis*, **53**, 4403-4416.



American Contingent Claims with Extra Information for the Buyer

Neda Esmaeeli¹, Peter Imkeller²

¹Sharif University of Technology, Department of Mathematical Sciences, Tehran, Iran

²Humboldt-Universität zu Berlin, Institut für Mathematik, Berlin, Germany

Abstract: We consider an American contingent claim on a financial market where the buyer has additional information. Both agents observe the same prices but their information may differ due to some extra information that has been obtained by the buyer in other fashion. The reference information flow is denoted by the filtration F and the buyer's information flow is modeled by an initial enlargement of the filtration F . The problem for the buyer is to choose a payoff from the class $\{R(\tau); \tau \in T\}$ which is optimal in the sense that it has a maximal expectation. As the buyer has more information, he has access to a larger set of available stopping times leading to a higher expected payoff. Therefore, it seems natural to investigate the value of the American contingent claim with asymmetric information. We provide a representation for this value in a suitable product space.

Keywords American contingent claims, Extra information, initial enlargements of filtrations.

Mathematics Subject Classification (2010): 60G40 60H30 62L15.

1 Introduction

A contingent claim is a contract that pays an amount $R(T)$ at time T which is called maturity or exercise time of the claim. The theory of contingent claim valuation in a complete market studied first by Black et al. (1973), Merton (1973, 1991), Harrison et al. (1979), Harrison et al. (1981), Duffie (1988), and Karatzas (1989). On the other hand, the pricing theory has been studied in the context of an incomplete market by Föllmer et al. (1990) and El Karoui et al. (1991). American contingent

¹Neda Esmaeeli : n.esmaili@mehr.sharif.ir

claims (ACCs), such as American call or put options, differ from their European counterparts in that they can be exercised by their holder at any time $0 \leq t \leq T$.

In our work, we consider an American contingent claim in which the buyer has more or superior information compared to the seller. While the seller makes his decisions according to public information flow $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$, the buyer possesses the additional information that comes from some random variable G and thus he accesses to the enlarged filtration $\mathbb{G} = (\mathcal{G}_t)_{t \in [0, T]}$. We suppose that \mathbb{G} is a version of the initial enlargement of \mathbb{F} by G which has usual conditions. We study the effect of this additional information on the value and the optimal exercise time of an American contingent claim. The situation is similar to the insider's optimal investment problem where an investor possessing some extra flow of information aims to maximize the expected utility on the final value of his portfolio. [Pikovsky et al \(1996\)](#) gave the first mathematical study of a utility maximization problem under additional initial information, posing the problem in the mathematical framework of an initially enlarged filtration. Further papers include [Elliott et al. \(1997\)](#), [Grorud et al. \(1998\)](#) and [Amendinger et al. \(1998\)](#). Building on results about initial enlargements of filtrations by [Jacod \(1985\)](#), we reduce the problem to a more suitable standard optimal stopping problem. To be more precise, we aim to use the density hypothesis to write the value function of an American contingent claim with extra information as the value of a new American contingent claim in the appropriate space.

2 Setup and Preliminaries

Let $T > 0$ represent a finite time. We consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$ is the reference filtration satisfying the usual conditions of right-continuity and completeness. Moreover we assume that \mathcal{F}_0 is trivial. We consider a random variable $G : \Omega \rightarrow \mathbb{R}$. Let $\bar{\mathbb{G}}$ be the initial enlargement of \mathbb{F} by G , i.e. $\bar{\mathbb{G}} = (\bar{\mathcal{G}}_t)_{t \in [0, T]}$ where $\bar{\mathcal{G}}_t = \mathcal{F}_t \vee \sigma(G)$, $t \in [0, T]$. Let finally $\mathbb{G} = (\mathcal{G}_t)_{t \in [0, T]}$ be the right-continuous regularization of $\bar{\mathbb{G}}$. We denote by P^G the law of G and for $t \in [0, T]$ by $P_t^G(\omega, du)$ the regular version of the conditional law of G given \mathcal{G}_t . Throughout this paper, we will assume that Jacod's density hypothesis ([Jacod \(1979\)](#), [Jacod \(1985\)](#)) stated in the following assumption is satisfied. For $t \in [0, T]$, the regular conditional law of G given \mathcal{G}_t is equivalent with the law of G for \mathbb{P} -almost all $\omega \in \Omega$.

i.e.

$$[G \in \cdot | _t] \sim (G \in \cdot), \text{ -a.s.}$$

According to [Jacod \(1985\)](#), for each $t \in [0, T]$ there exists an $\otimes(\cdot)$ -measurable version of $\alpha_t(u)(\omega) := \frac{dP_t^G(u, \omega)}{dP^G(u)}$ which is strictly positive. And for each $u \in \mathbb{H}$, $\{\alpha_t(u)\}_{t \in [0, T]}$ is a martingale w.r.t. \mathbb{H} . Let $t \in \mathbb{H}^+$ and \mathbb{H} a filtration in \mathcal{F} . We denote by $\mathcal{T}_{t, T}(\mathbb{H})$ the set of \mathbb{H} -stopping times with values in $[t, T]$.

Definition 2.1. *Consider the following payoff process*

$$R = L1_{[0, T)} + \xi 1_{\{T\}}, \quad (2.1)$$

where L is an \mathbb{H} -adapted real-valued càdlàg process and ξ an T -measurable random variable, satisfying the integrability condition

$$\left[\sup_{t \in [0, T]} |L_t| + |\xi| \right] < \infty. \quad (2.2)$$

For $t \in [0, T]$, $\tau \in \mathcal{T}_{t, T}(\mathbb{H})$, the value function of an American contingent claim is defined by

$$V(t) = \sup_{\tau \in \mathcal{T}_{t, T}(\mathbb{H})} [R(\tau) | _t]. \quad (2.3)$$

τ is the buyer's stopping time and plays the role of a control tool. We suppose throughout this paper that $0 \leq L_T \leq \xi < +\infty$. We need to introduce some further notations.

Consider the probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$, where

$$\begin{aligned} \widehat{\Omega} &:= \Omega \times \mathbb{H}, \\ \widehat{\mathcal{F}}_t &:= \bigcap_{s > t} (s \otimes \mathcal{B}(\mathbb{H})), \quad t \in [0, T], \\ \widehat{\mathbb{P}}_t &:= (\mathbb{P}_t)_{t \in [0, T]}, \quad \widehat{\mathbb{P}} = \otimes \mathcal{B}(\mathbb{H}), \\ \widehat{\eta} &:= \otimes \eta, \end{aligned} \quad (2.4)$$

where η is a probability measure on $(\mathbb{H}, \mathcal{B}(\mathbb{H}))$. We denote by $\widehat{\mathbb{E}}_t$ the expectation w.r.t. $\widehat{\mathbb{P}}_t$.

3 American contingent claims in an initially enlarged filtration

We consider an American contingent claim where, in contrast to the seller, the buyer possesses additional information. This extra information may be based for instance on a good analyst or better software. The additional information is described by the random variable we denote by G . A natural question one may ask is "what is the value of an American contingent claim with extra information?"

A filtration usually encodes a flow of information. So it is natural to model extra information by an enlargement of a filtration. We will consider an initial enlargement of the reference filtration. This means that we add all the extra information at initial time to the reference filtration. As introduced above, $\mathbb{G} = (\mathcal{G}_t)_{t \in [0, T]}$ is the initial enlargement of \mathbb{F} by G . Due to the definition of the value function of an American contingent claim (??), our first step on the way to answer the above questions is to study

$$\tau \in \mathcal{T}_{t, T}(\mathbb{G}) [R(\tau)|_t], \quad (3.1)$$

where $\mathcal{T}_t = \mathcal{T}_t$. We also study the case $\mathcal{T}_t = \mathcal{T}_t$ which will be seen to be understood as an optimal projection problem. Our main idea is to look for a suitable representation of \mathbb{G} -stopping times and then reduce the problem to a corresponding problem in a product filtration which contains the reference filtration. The following result gives a useful clue to calculate conditional expectations with respect to the larger filtration.

Lemma 3.1. *Suppose that $X : \widehat{\Omega} \times [0, T] \rightarrow \mathbb{R}$ is a process, $t \in [0, T]$ and $G : \Omega \rightarrow \mathbb{R}$ a random variable such that $X_t(G)$ is \mathcal{G}_t -measurable and \mathcal{G}_t -integrable. Then for $s \leq t$*

$$[X_t(G)|_s] = \frac{1}{\alpha_s(G)} [X_t(u)\alpha_t(u)|_s]_{u=G}.$$

Proof. See [Callegaro \(2013\)](#), p. 5. □

Theorem 3.2. *Let $t \in [0, T]$. Under Assumption 1.1 on G and the integrability condition (2.2) on R ,*

we have for $t \in [0, T]$

$$V_t^G :=_{\tau \in t, T(\mathbb{G})} [R(\tau)|_t] = \frac{1}{\alpha_t(G)} \left(\tau(\cdot) \in t, T \cap \widehat{[R(\cdot, \tau(\cdot))|_t]} \right)_G.$$

References

- Amendinger, J., Imkeller, P. and Schweizer, M. (1998). Additional logarithmic utility of an insider, *Stochastic Processes and their Applications*, **75**, 263-268.
- Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **3**, 637-654.
- Callegaro, G., Jeanblanc, M. and Zargari, B. Carthaginian enlargement of filtrations, *ESAIM: Probability and Statistics*, **17**, 550-566.
- Duffie, D. (1988). *Security markets: Stochastic models*, Boston, Academic Press.
- Elliott, R.J., Geman, H. and Korkie, B.M. (1997). Portfolio optimization and contingent claim pricing with differential information, *Stochastic and Stochastic Reports*, **60**, 185-203.
- El Karoui, N., Quenez, M.C. Programmation dynamique et évaluation des actifs contingents en marché incomplet, *Comptes Rendus de l'Académie des Sciences. Paris, Sér.*, **1**, 851-854.
- Föllmer, H., Schweizer, M. (1990). *Hedging of contingent claims under incomplete information*, Applied Stochastic Analysis, eds. M. H. A. Davis and R. J. Elliot. London: Gordon and Breach.
- Gorud, A., Pontier, M. (1998). Insider trading in a continuous time market model, *International Journal of Theoretical and Applied Finance*, **1**, 331-347.
- Gorud, A., Pontier, M. (1998) Asymmetric information and incomplete market, *International Journal of Theoretical and Applied Finance*, **4**, 285-302.

- Harrison, M., Kreps, D. (1979). Martingale and arbitrage in multiperiod securities markets, *Journal of Economic Theory*, **20**, 381-408.
- Harrison, M., Pliska, S.R. (1981). Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes and their Applications*, **11**, 215-260.
- Jacod, J. (1979). *Calcul stochastique et problèmes de martingales*, Lecture Notes in Mathematics 714. Springer-Verlag, Berlin.
- Jacod, J. (1985). *Grossissement Initial, Hypothèse(H') et Théorème de Girsanov*, Lecture Notes in Mathematics 1118. Springer-Verlag, Berlin.
- Karatzas, I. (1989). Optimization problems in the theory of continuous trading, *SIAM. Journal on Control and Optimization*, **27**, 1221-1259.
- Kobylanski, M., Quenez, M.C. (2012). Optimal stopping time problem in a general framework, *Electronic Journal of Probability*, **17**, 72, 1-28.
- Merton, R. Theory of rational option pricing, *Bell Journal of Economics and Management Sciences*, **4**, 141-183.
- Merton, R. (1991). *Continuous time finance*, Basil Blackwell
- Pikovsky, I., Karatzas, I. (1996). Anticipative portfolio optimization, *Advances in Applied Probability*, **28**, 1095-1122.



Reliability analysis of a semi-Markov repairable system

Malek Fathizadeh¹, Omid Kharazmi

Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.

Abstract: In this paper, we attempt to present the applications of Markov renewal matrix in reliability analysis of repairable systems. In this regard, a semi-Markov process with $n-k+2$ states is applied to model the system. A recursive formula is proposed for obtaining the associated Markov renewal matrix. Several important reliability measures such as expected number of working components, expect number of system recoveries, reliability function, mean time to first failure, mean time between failures, and etc. are obtained using the Markov renewal matrix.

Keywords Laplace transform, Markov renewal matrix, Mean time to failure, Semi-Markov processes, Tridiagonal matrix.

Mathematics Subject Classification (2010): 60K15 90B25.

1 Introduction

In recent decades, k -out-of- n systems are encountered with many industrial applications such as electronics industry, telecommunication network systems, power generator and transmission systems. For this reason, k -out-of- n systems with various configurations have attracted substantial research interest in reliability for example see Akhtar (1994), Erilmaz (2009) and Moustafa (1996). Among these researches, the reliability analysis of repairable k -out-of- n systems have been received special attentions. Exponential distribution is one of the most commonly considered assumptions for lifetimes in this type of systems. Under this assumption, many of authors have employed Markov processes, and have obtained little reliability indices such as mean time to failure and reliability function for repairable k -out-of- n systems see Li et al. (2006) and Moustafa (1996). The issue that has so far received less

¹Malek Fathizadeh : m.fathi@vru.ac.ir

attention is using the semi-Markov process in the analysis of such systems. The potentiality of semi-Markov processes for the analysis of reliability systems is illustrated by many of authors Bieth et al. (2009), Li and Yam (2006), Perez and Castro (2002) and Yin et al. (2002). Recently, Fathizadeh and Khorshidian (2015) have proposed an alternative approach to the analysis of semi-Markov systems. They considered a two unit cold standby system with imperfect switch as a three state semi-Markov process, and obtained several important reliability measures such as availability, expected number of failures, expected number of visits and mean up and down times of the system by means of the associated Markov renewal matrix.

In this paper, a k-out-of-n: G system with a repairman is considered as $n - k + 2$ states semi-Markov model. A recursive formula is proposed for obtaining the associated Markov renewal matrix. Using Markov renewal matrix, the collection of important reliability measures include expected number of working components, expect number of system recoveries, reliability function, mean time to failure, mean time between failures, expected number in which number of working components reduces to k during time interval 0 to first failure and etc. are obtained.

The paper is organized as follows. In Section 2 some basic notation and definitions concerning semi-Markov processes are given. In Section 3 several important reliability indices has been obtained for k-out-of-n: G system under a semi-Markov setting.

2 Preliminaries

In this section we provide brief preliminaries on semi-Markov processes. For more on semi-Markov processes and their applications we refer readers to Cinlar (1975), and Osaki (1992). A Markov renewal process may be defined as follows:

Definition 2.1. *Let \mathbb{L} be a countable set and $(\mathbf{J}, \mathbf{T}) = \{(J_n, T_n), n = 0, 1, 2, \dots\}$ denote a two-dimensional stochastic process with values in $\mathbb{L} \times [0, \infty)$, then (\mathbf{J}, \mathbf{T}) is a Markov renewal process if for all $i, j \in \mathbb{L}$,*

$$P\{J_{n+1} = j, T_{n+1} - T_n \leq t | J_n, T_n, \dots, J_0, T_0\} = P\{J_{n+1} = j, T_{n+1} - T_n \leq t | J_n\}.$$

We assume that the state space \mathbb{L} is finite, the process is time homogenous with transition prob-

abilities $\varrho_{ij}(t) = P\{J_{n+1} = j, T_{n+1} - T_n \leq t | J_n = i, j \in \mathbb{L}, n \geq 0$ and $\mathbf{Q}(t) = [Q_{ij}(t)]_{i,j \in \mathbb{L}}, t \in \mathbb{R}^+$, as the matrix of renewal kernels. Note that $\{J_n = 0, 1, \dots\}$ and $\{T_n = 0, 1, \dots\}$ are the state and time instant of n -th renewal. The associated counting process representing the total number of transitions within $[0, t]$ is denoted by $\{N(t) : t \geq 0\}$.

Definition 2.2. A stochastic process $\{X(t) : t \geq 0\}$ is called a semi-Markov process on \mathbb{L} , generated by Markov renewal process (\mathbf{J}, \mathbf{T}) , with kernel $\mathbf{Q}(t), t \geq 0$, if $X(t) = J_{N(t)}$.

Since the trajectory of the semi-Markov process has constant value on the half-interval $[T_n, T_{n+1}]$ and is continuous from the right, $X(t) = X(T_n) = J_n$, for $t \in [T_n, T_{n+1}]$. We denote the Markov renewal matrix by $\mathbf{R}(t) = [R_{ij}(t)]_{i,j \in \mathbb{L}}$, where $R_{ij}(t) = E[N_{i,j}(t)]$, with $N_{i,j}(t)$ as the number of visits to the state j within $[0, t]$ given $X(0) = i$. In the case of finite \mathbb{L} , it is well known that

$$\bar{\mathbf{R}}(s) = (I - \bar{\mathbf{Q}}(s))^{-1} \quad (2.1)$$

where $\bar{\mathbf{R}}(s)$ and $\bar{\mathbf{Q}}(s)$ are the Laplace-Stieltjes transform of $\mathbf{R}(t)$ and $\mathbf{Q}(t)$, respectively. Using Markov renewal matrix we can obtain $P(X(t) = j | X(0) = j)$ in the following way:

$$P_{ij}(t) = \int_0^t R_{ij}'(u) h_j(t-u) du + \delta_{ij} h_j(t) R_{ij}(0) \quad (2.2)$$

where

$$h_i(t) = P\{T_1 > t | X(0) = i\} = 1 - \sum_{j \in \mathbb{L}} Q_{ij}(t), i \in \mathbb{L}, \delta_{ij} = 0, i \neq j, \delta_{ij} = 1, i = j.$$

We denote first passage time distribution from state i to state j by $G_{ij}(t) = P(N_{ij}(t) > 0)$. Let us to define the random variable $T_1(i|j), n \geq 1$ as the times of first returns to state j with starting from i as initial state. So, it is easy to known that $G_{ij}(t) = P(T_1(i|j) < t)$.

For the Laplace transform of the first return time distributions, we have:

$$\bar{G}_{ij}(s) = \frac{\bar{R}_{ij}(s)}{1 + \bar{R}_{ij}(s)}, i, j \in \mathbb{L} \quad (2.3)$$

Definition 2.3. A matrix $\mathbf{A} = [A_{ij}]$ is tridiagonal if $a_{ij} = 0$ whenever $|i - j| > 1$.

Let $a_i = a_{ii}$ for $i = 1, \dots, n$, let $b_i = a_{i,i+1}$ and $c_i = a_{i+1,i}$ for $i = 1, \dots, n - 1$. Then the tridiagonal

matrix takes the form

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 & 0 & \dots & 0 & 0 & 0 \\ c_1 & a_2 & b_2 & \dots & 0 & 0 & 0 \\ 0 & c_2 & a_3 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & a_{n-2} & b_{n-2} & 0 \\ 0 & 0 & 0 & \dots & c_{n-2} & a_{n-1} & b_{n-1} \\ 0 & 0 & 0 & \dots & 0 & c_{n-1} & a_n \end{pmatrix}$$

Let $\mathbf{A}_k = [A_{ij}]$, $i, j = 1, \dots, k$. The determinant of \mathbf{A} can be calculated starting with $\det \mathbf{A}_1 = a_1$, $\det \mathbf{A}_2 = a_1 a_2 - b_1 c_1$, and then computing the following recursive relation

$$\det \mathbf{A}_{k+1} = a_{k+1} \det \mathbf{A}_k - b_k c_k \det \mathbf{A}_{k-1}, \quad k = 2, \dots, n-1, \quad (2.4)$$

Horn and Johnson (2013).

3 Reliability analysis of k-out-of-n G:system

In this section, a k-out-of-n G:system with a repairman is modeled by a semi-Markov process and several important reliability measures of the semi-Markov system is derived using the associated Markov renewal matrix.

3.1 Semi-Markov system and the associated Markov renewal matrix

The systems investigated here have the following characteristics:

- The system consists of n components. X_1, X_2, \dots, X_n are independent and identical component lifetimes with $F_X(x) = 1 - e^{-\lambda x}$, and $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ denote the order statistics corresponding to X_1, X_2, \dots, X_n .
- There is a single repair facility with independent and identically repair times with $G_Y(y) = 1 - e^{-\mu y}$.
- The system is working if at least k components are working. If $n - k + 1$ components failed, the system restarts working as soon as one of the failed components is repaired.
- All random variables are assumed to be mutually independent.

To describe the model and derive the associated renewal kernel, we will introduce the following states:

0: All components are working,

i : i components are failed ($i = 1, 2, \dots, n - k$).

$n-k+1$: the system is failed .

Let the process $X(t), t > 0$ the states of the system at time t , with the state space $\mathbb{L} = \{0, 1, \dots, n-k+1\}$.

According to definition 1 and 2 it is easy to know that $X(t), t > 0$ is semi-Markov process with following kernels:

$$\mathbf{Q}(t) = \begin{pmatrix} 0 & Q_{01}(t) & \cdots & 0 & 0 \\ Q_{10}(t) & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 0 & Q_{N-K, N-K+1}(t) \\ 0 & 0 & \cdots & Q_{N-K+1, N-K}(t) & 0 \end{pmatrix}$$

where

$$Q_{01}(t) = P(X_{1:n} < t) = 1 - e^{-\lambda t}$$

$$Q_{ij}(t) = P(X_{i+1:n} < Y, X_{i+1:n} < t) = \sum_{r=0}^i \left(\frac{n!}{i!(n-i-1)!} (-1)^r C(i, r) \right) \\ \times \frac{\lambda(1 - e^{-(\lambda+(n-i-1)\lambda+r\lambda+\mu)t})}{\lambda + (n-i-1)\lambda + r\lambda + \mu} \quad i - j = -1 \text{ and } i \neq 0,$$

$$Q_{ij}(t) = P(Y < X_{i+1:n}, Y < t) = \sum_{r=0}^i \left(\frac{n!}{i!(n-i-1)!} (-1)^r C(i, r) \right) \\ \times \frac{\lambda\mu(1 - e^{-(\lambda+(n-i-1)\lambda+r\lambda+\mu)t})}{(\lambda + (n-i-1)\lambda + r\lambda + \mu)(\lambda + (n-i-1)\lambda + r\lambda + \mu)} \quad i - j = 1 \text{ and } i \neq n - k + 1,$$

$$Q_{n-k+1, n-k}(t) = P(Y < t) = 1 - e^{-\mu t}.$$

And therefore

$$(I - \overline{\mathbf{Q}}(s))^{-1} = [M_{ij}(s)]^{-1}.$$

where,

$$M_{01}(s) = -\frac{n\lambda}{s + n\lambda},$$

$$M_{ij}(s) = -\sum_{r=0}^i \left(\frac{n!}{i!(n-i-1)!} (-1)^r C(i, r) \frac{\lambda}{\lambda + (n-i-1)\lambda + r\lambda + \mu + s} \right) \quad i - j = -1 \text{ and } i \neq 0,$$

$$M_{ij}(s) = -\sum_{r=0}^i \left(\frac{n!}{i!(n-i-1)!} (-1)^r C(i, r) \right) \times \frac{\lambda}{(\lambda + (n-i-1)\lambda + r\lambda + \mu + s)(\lambda + (n-i-1)\lambda + r\lambda + \mu)} \quad i - j = 1 \text{ and } i \neq n - k + 1,$$

$$M_{n-k+1, n-k}(s) = -\frac{\mu}{\mu + s},$$

$$M_{ij}(s) = 1 \quad i = j.$$

Inverse of the above tridiagonal matrix can be calculated inductively by using the relation (1.4). By using the relation (1.1) and then taking inverse Laplace transform of $\bar{\mathbf{R}}(s)$ we can reach to the Markov renewal matrix. For more insight about Markov renewal Matrix calculations see Fathizadeh and Khorshidian (2015). After calculating Markov renewal Matrix and then using relations (1.2) and (1.3) we can reach to the semi-Markov transition probability matrix and first passage time distributions.

3.2 Reliability measures

In this subsection we try to do a comprehensive study on the semi-Markov system. It is worth noticing that in the all obtained measures Markov renewal matrix plays a fundamental role.

- 1) Reliability function for the system is $1 - G_{0, n-k+1}(t)$.
- 2) Availability of the system is $1 - P_{0, n-k+1}(t)$.
- 3) Mean time to system failure is $\int_0^\infty u dG_{0, n-k+1}(u)$.
- 4) Mean time between failures of the system is $\int_0^\infty u dG_{n-k+1, n-k+1}(u) - \mu$.
- 5) Mean down time during time interval $(0, t]$ is $\int_0^t P_{0, n-k+1}(u) du$ and consequently, mean up time is $t - \int_0^t P_{0, n-k+1}(u) du$.
- 6) Busy period of the repairman during time interval $(0, t]$ is $1 - \int_0^t P_{00}(u) du$.
- 7) Expected number of failure of the whole system during time interval $(0, t]$ is $R_{0, n-k+1}(t)$.
- 8) Expected number of working components at time t is $\sum_0^{n-k} (n-i) P_{0i}(t)$.

References

- Akhtar, S., Chandrasekar, B. and Balakrishnan, N. (2008). Reliability of k-out-of-n: G Systems with imperfect faultcoverage, *IEEE Transactions on Reliability*, **43(1)**, 101–106.
- Bieth, B., Hong, L. and Sarkar, J. (2009). A standby system with two types of repair persons, *Applied Stochastic Models in Business and Industry*, **26**, 577–594.
- Cinlar, E. (1975). *Introduction to Stochastic Processes*, Prentice-Hall.
- Erilmaz, S. (2010). Number of working components in consecutive k-out-of-n system while it is working, *Communications in Statistics–Simulation and Computation*, **39**, 383–392.
- Fathizadeh, M. and Khorshidian, K. (2015). An alternative approach to reliability analysis of cold standby systems, *Communications in Statistics–Theory and Methods* .**29**, 341–352.
- Horn, R.A. and Johnson, C.R. (2013). *Matrix Analysis*, 2nd ed, Cambridge University Press, New York.
- Li, X, Zou, M. J and Yam, R. C. (2006). Reliability analysis of a repairable k-outof-n system with some components being suspended when the system is down, *Reliability Engineering and System Safety*, **95**, 305–311.
- Moustafa, M. S. (1996). Transient analysis of reliability with and without repair for k-out-of-n:G systems with two failure modes, *Reliability Engineering and System Safety*, **53**, 31–35.
- Osaki, S. (1992). *Applied Stochastic System Modeling*, Springer, Berlin .
- Perez , R. and Castro, I. (2002). A reliability semi-Markov model involving geometric processes, *Applied Stochastic Models in Business and Industry*, **18**, 157–170.
- Yin, L., Fricks, R. M. and Trivedi, K. S. (2002). Application of semi-Markov process and CTMC to evaluation of UPS system availability, *Proceedings Annual Reliability and Maintainability Symposium*, 584–591.



Improving in hybrid Monte Carlo via Halton sequence with increasing and decreasing informity

Behrouz Fathi-Vajargah ¹, Hamzeh Torabi ², Hojatollah Zakerzadeh ³

Ali.A.L.Zadeh ⁴

¹Department of Statistics, University of Guilan, Rasht, Iran

^{2,3,4} Department of Statistics, University of Yazd, Yazd, Iran

Abstract: The Monte Carlo is a good numerical method in estimation problems. In this paper we apply a new method by using MC method. This method is consisting of three steps. First step is selection number of sub intervals, in second step generating random numbers of uniform (0,1) distribution and third step we must conduct them to k sub interval (such that k is the number of optional sub intervals on (0,1)). Additional we use hybrid Monte Carlo for integral estimation. It is consist of Halton sequence and Monte Carlo or Halton sequence and partition Monte Carlo (PMC (is new method)).

Keywords low-discrepancy sequences, Halton sequence, Partition Monte Carlo, hybrid sequences, quasi-Monte Carlo integration.

Mathematics Subject Classification (2010): 11K45.

1 Introduction

Quasi-Monte Carlo (QMC) methods are widely used for high-dimensional integrals. Quasi-Monte Carlo sequences such as [Niederreitr \(1992\)](#) and Halton are popular class of low-discrepancy sequences, and Halton sequence was introduced by [Halton \(1960\)](#). simplest hybrid sequence has been combined of random sequence and low-discrepancy sequence, some of researchers worked in this issue, [Caffisch et al. \(1996\)](#); [Matousek \(1998\)](#); [Morokofft William et al. \(1993\)](#).

⁴Speaker's Name : Ali.A.L.zadeh@stu.yazd.ac.ir

2 Preliminary in Halton sequence

Monte Carlo method is an analytical technique for solving a problem by performing a large number of trial runs, called simulations, and inferring a solution from the collective results of the trial runs. Monte Carlo integration is to use random points for the numerical evaluation of an integral [Zdravko et al. \(2011\)](#), $I = \int_a^b f(x)dx \approx \frac{b-a}{N} \sum_{i=1}^N f(X_i)$, in this method I was approximated by taking random variables X_i and arithmetic averaging by contribution $f(X_i)$. The base of Monte Carlo method is to generate i.i.d. random variables on $(0, 1)$, which the interval number uniformly distributed on $[0, 1)$ to generate these distributed random variables, we employ the rand function on Matlab programming software. However we significantly improved our the results by managing and partitioning of the random numbers generated that is known as PMC method [Fathi et al. \(2014\)](#).

Monte Carlo method is used for integrals estimating, solving linear equations, finding eigenvalues of the matrix and finance engineering and etc., have revealed some demerits including lack of efficiency, low convergence speed, high radius of convergence. So, after so many achievements in this issue, eventually the Quasi-Monte Carlo methods has been proposed instead of the Monte Carlo method for covering all of the before mentioned inefficiencies. Briefly, in QMC the following relation is hold;

$$I = \int_{I^s} f(U)dU \approx \frac{1}{N} \sum_{i=1}^N f(x_i)$$

QMC methods can be viewed as deterministic versions of Monte Carlo methods. Quasi- Monte Carlo methods are based on quasi-random sequences, and the best sequences and most commonly used of these sequences are: Halton, Sobol, Faure and Niederreiter sequences [Chi \(2004\)](#); [Halton \(1960\)](#); [Niederreitr \(1992\)](#). In this paper we use the following Halton sequence.

For an integer $b \geq 2$, let $Z_b = \{0, 1, \dots, b-1\}$ denote the least residue system modulo b . Let $n = \sum_{j=1}^{\infty} a_j(n)b^{j-1}$ with all $a_j(n) \in Z_b$ and $a_j(n) = 0$ for all sufficiently large j be the unique digit expansion of the integer $n \geq 0$ in base b , [Fox \(1986\)](#); [Zdravko et al. \(2011\)](#). The radical-inverse function in base b_j for $j = 1, 2, \dots, s$ is defined by $\phi_{b_j}(n) = \sum_i \sigma(a_i(j, n)b^{-i-1})$ Where σ is the permutations on $a_i(j, n)$'s. For pairwise coprime integers $b_1, \dots, b_s \geq 2$, the Halton sequence in the bases b_1, \dots, b_s is given by $X_n = (\phi_{b_1}(n), \dots, \phi_{b_s}(n)) \in [0, 1)^s$ for $n = 0, 1, \dots$ [Halton \(1960\)](#).

We remind that quasi-random sequence for high dimension will lost their own effectiveness; in other

word the high dimensions correlation of these points will be increased and concentration will be placed around the lines, therefore for making up this demerit these points will be scrambled. [Kocis et al. \(1997\)](#), and this method called "RR2".

If f has bounded variation $V_{HK}(f)$ on $[0, 1]^s$ in the sense of Hardy and Krause, then, for any $X_1, \dots, X_N \in [0, 1]^s$ then, we have $\left| \frac{1}{N} \sum_{i=1}^N f(X_i) - I(f) \right| \leq V_{HK}(f) \cdot D_N^*(X_1, \dots, X_N)$, Where $D_N^*(X_1, \dots, X_N)$ is the star-discrepancy of point set $\{X_1, \dots, X_N\}$. The star discrepancy of the first n Halton points in dimension s with relatively prime bases b_1, \dots, b_s in $D_N^*(X) \leq C_s \frac{1}{N} (\log N)^s + O(\frac{1}{N} (\log N)^{s-1})$ that C_s is coefficient for the Halton [Chi \(2004\)](#); [Domingo et al. \(2013\)](#); [Okten \(2001\)](#), $C_s = \prod_{j=1}^s \frac{p_j - 1}{2 \log p_j}$.

In order to minimize error, we should be reduce $D_N^*(X)$ or $V(f)$.

3 Hybrid sequences

This mixed method was made up mixed sequence. Mixed sequence is combined MC sequence (or PMC sequence) by randomized (scrambled) QMC sequence (RQMC), [Okten \(2006\)](#). Padding PMC sequence by randomized QMC (RQMC). We are consider the problem of estimate the multidimensional integration

$$I = I(f) = \int_{I^s} f(X) dX \int_0^1 \dots \int_0^1 f(X_1, \dots, X_s) dX_1 \dots dX_s$$

according to the Monte Carlo method [Zdravko et al. \(2011\)](#), $\hat{I} = \frac{1}{N} \sum_{k=1}^N f(X^{(k)})$, where $X^{(k)}$ are s -dimensional vectors chosen adaptable, where $x^{(k)}$ for $k = 1, 2, \dots, N$, are sequences of variables s -dimensional. Let (x_1, \dots, x_d) be d -dimensional subset of variables (x_1, \dots, x_s) , for $d \leq s$. Then one has the following points [Okten \(2006\)](#):

1. Sample (x_1, \dots, x_d) using a d -dimensional MC sequence and for (x_{d+1}, \dots, x_s) of the variables use $(s - d)$ -dimensional low-discrepancy sequence (padding MC by QMC);
2. Sample (x_1, \dots, x_d) using a d -dimensional partition Monte Carlo sequence and for (x_{d+1}, \dots, x_s) of the variables use $(s - d)$ -dimensional low-discrepancy sequence (padding PMC by QMC);

3. Sample (x_1, \dots, x_d) using a d -dimensional MC sequence and for (x_{d+1}, \dots, x_s) of the variables use $(s-d)$ -dimensional RQMC sequence (padding MC by RQMC);
4. Sample (x_1, \dots, x_d) using a d -dimensional Partition Monte Carlo sequence and for (x_{d+1}, \dots, x_s) of the variables use $(s-d)$ -dimensional RQMC sequence (padding PMC by RQMC).

Here $x^{(k)} = (X^{(k)}, q^{(k)})$ be a s -dimensional sequence by concatenating by vectors $X^{(k)}$ and $q^{(k)}$. In the first strategy, $X^{(k)}$, $k \geq 1$, are independent random variables with the uniform distribution on $(0, 1)^d$, and $(q^{(k)})_{k \geq 1}$ is a $(s-d)$ -dimensional low-discrepancy sequence. In the second strategy, $X^{(k)}$, $k \geq 1$ are partitioned independent random variables with the uniform distribution on $(0, 1)^d$, and $(q^{(k)})_{k \geq 1}$ is similar to the first strategy. In the third strategy, $X^{(k)}$, $k \geq 1$ are independent random variables with the uniform distribution on $(0, 1)^d$, and $(q^{(k)})_{k \geq 1}$ is a $(s-d)$ -dimensional randomized Quasi-Monte Carlo (RQMC) sequence. Last one, $X^{(k)}$, $k \geq 1$ are similar to the second strategy, and $(q^{(k)})_{k \geq 1}$ is similar to the third strategy. All of the above defined sequences are called mixed sequences.

Theorem 3.1. *Let $x^{(k)} = (X^{(k)}, q^{(k)})$ be a s -dimensional mixed sequence, where $\{q^{(k)}\}_{k=1}^{\infty}$ be a $(s-d)$ -dimensional low-discrepancy sequence with $D_N^*(q^{(k)}) \leq C_{(s-d)} \frac{1}{N} (\log N)^{s-d} + O(\frac{1}{N} (\log N)^{s-d-1})$ and $X^{(k)}$ is a random variable with the uniform distribution on $(0, 1)^d$ then, for sufficiently large N , and for $0 \leq a \leq 1$ and the star discrepancy for mixed (s, d) sequence satisfies [Okten \(2001\)](#),*

$$D_N^*(x^{(k)}) \prec \frac{1}{N^{a/2}} + C_{(s-d)} \frac{1}{N} (\log N)^{s-d} + O(\frac{1}{N} (\log N)^{s-d-1})$$

this star discrepancy with probability greater than or equal to $1 - 2e^{-2N^{1-a}}$. In other words

$$P(D_N^*(x^{(k)}) \prec \frac{1}{N^{a/2}} + C_{(s-d)} \frac{1}{N} (\log N)^{s-d} + O(\frac{1}{N} (\log N)^{s-d-1})) \geq 1 - 2e^{-2N^{1-a}}$$

proof in [Okten \(2006\)](#).

In the table, we compute D_1^* and D_2^* , that the D_1^* is upper bound for the discrepancy of the s -dimensional Halton sequence, and D_2^* be the probabilistic upper bound for the corresponding mixed (s, d) sequence. The best values for C_d , $2 \leq d \leq 10$ of the Halton sequence, and corresponding mixed (s, d) , sequence. In this table using three digits rounding, and constant values $N = 10^7$, $a = 0.99$, $d = s/2$ and $s = 4, 6, \dots, 10$. The numerical results are becoming clearer for them in drawing together

Table 1: Bounds for the discrepancy.

s	d	D_1^*	D_2^*
4	2	5.167×10^{-3}	3.655×10^{-4}
6	3	3.270×10^{-2}	6.855×10^{-4}
8	4	$2.670 \times 10^{+2}$	5.51×10^{-3}
10	5	$6.070 \times 10^{+4}$	7.825×10^{-2}

the team apart from the figures.

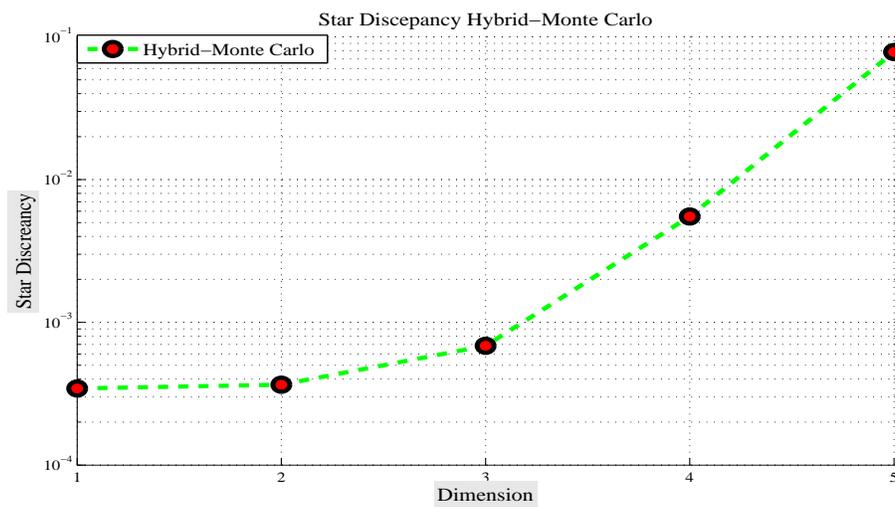


Figure 1: Star discrepancy for hybrid sequence padding MC by Halton.

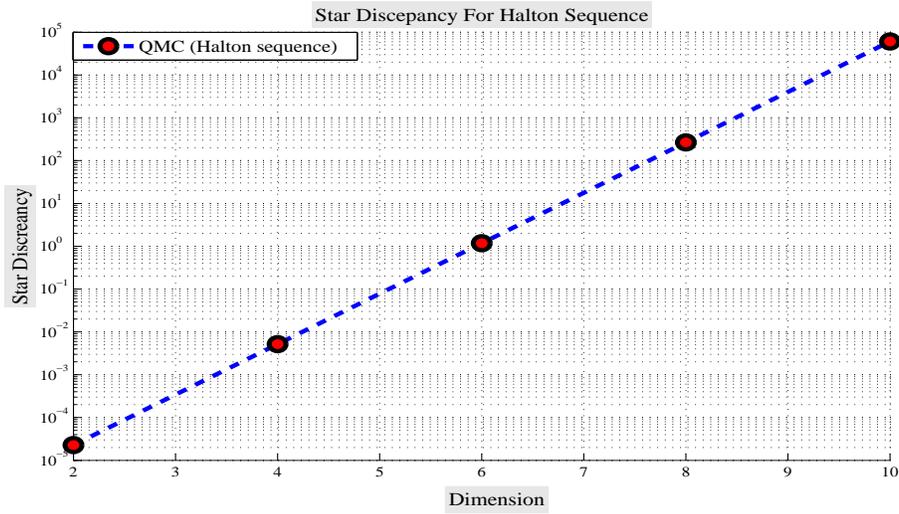


Figure 2: Star discrepancy for Halton sequence.

The efficiency of the hybrid sequence is illustrated by the following figures.

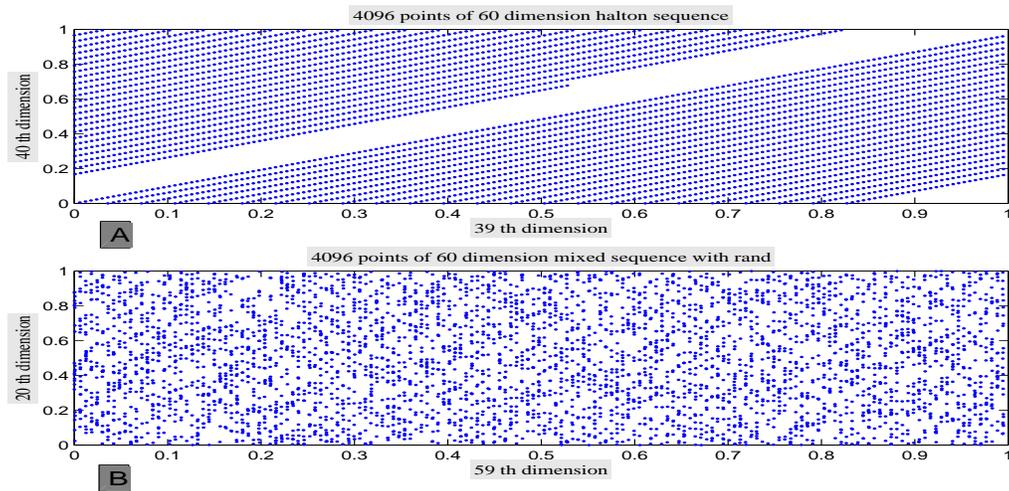


Figure 3: Fig 3.A ; represents the point sets for 60-dimensional Halton sequence which stand for bases of 167 and 173, and in fig 3.B; X-axis Halton sequence for base of 167 and Y-axis random numbers generated(N=4096).

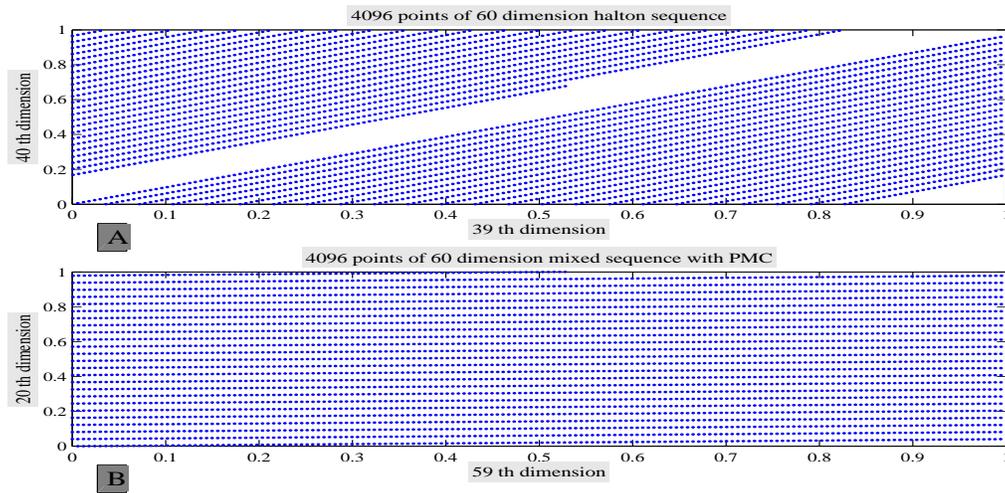


Figure 4: Fig 4.A; represents the point sets for 60-dimensional Halton sequence which stand for bases of 167 and 173, and fig4. B; is allotted to the point sets for 60-dimensional mixed sequence, with X-axis Halton sequence for base of 167 and Y-axis partitioning random numbers generated (N=4096).

4 Numerical results

In this section to testify the efficiency of the PMC sequence the following integral will be examined. According to the study it is concluded that if the dimension of scramble or deterministic section be equal 3, then the integral estimation error would be the lowest one. $I = \int_0^1 \dots \int_0^1 \prod_{i=1}^s |4x_i - 2| dx_1 \dots dx_s = 1$ We estimated this integral in the case of $s = 50$ and $s = 70$.

References

Caffisch, R., Moskowitz, B., (1996). Smoothness and dimension reduction in quasi-Monte Carlo methods, *Math. Comput. Modelling* 23 pp. 37–54.

Chi, H. (2004). Scrambled quasirandom sequences and their application, *The florida state university, college of arts and science.*,

Domingo, G., Roswitha, H. and Niederreiter, H. (2013). A general discrepancy bound for hybrid sequences involve Halton sequence. *Uniform Distribution Theory*8, pp. 31–45.

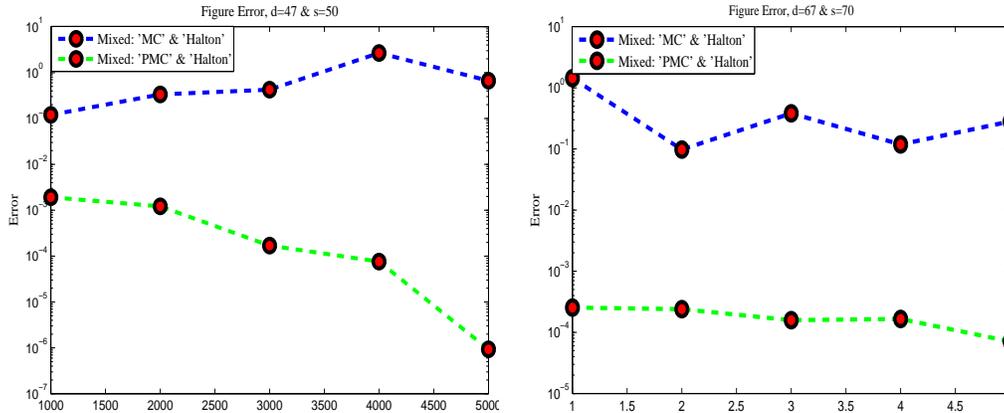


Figure 5: Integral estimation error I_N by use of mixed sequence, padding MC by Halton in blue curve and padding PMC by Halton in green curve, where $s = 50$ (left fig) and $s = 70$ (right fig).

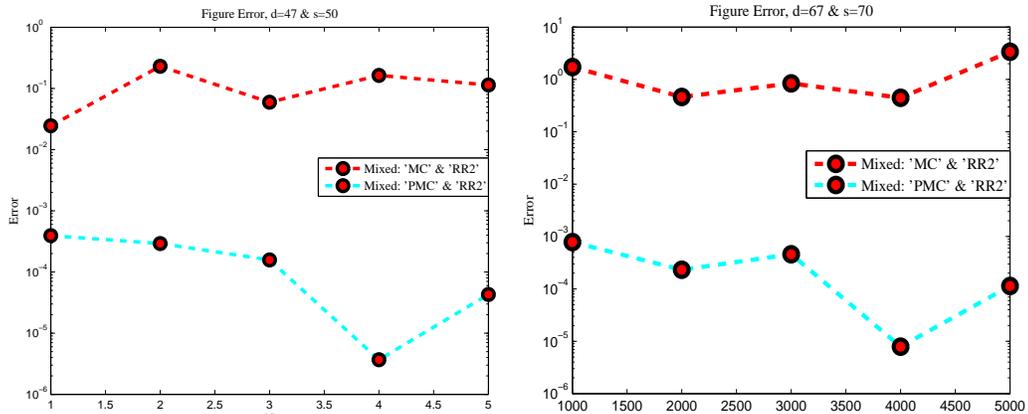


Figure 6: Integral estimation error I_N by use of randomized mixed sequence, padding MC by RQMC(RR2) in red curve and padding PMC by RQMC(RR2) in cyan curve, where $s = 50$ (left fig) and $s = 70$ (right fig).

Fathi-Vajargah, B., L-Zadeh, Ali, A., (2014). On the efficiency of partition pseudo random number generated in integral estimation, *Engineering Mathematics Letters*, ISSN: 2049-9337.

Fox, B. (1986). Implementation and relative efficiency of quasirandom sequence generators. *ACM Trans. on Mathematical Software*, p. 362–376.

- Halton, J. (1960). On the efficiency of Certain Quasirandom Sequences of Points in Evaluating Multi Integrals, *Numerisce Mathematic*, pp. 84–90.
- Kocis, L., and W. J. Whiten(1997). "Computational Investigations of Low-Discrepancy Sequences.*ACM Transactions on Mathematical Software. Vol. 23, No. 2*, pp. 266–294.
- Krykova, I. (2003). Thesis, Evaluating of path-dependent securities with low discrepancy methods. *Faculty of the Worcester polytechnic institute*.
- Matousek, J. (1998), On the L2-discrepancy for anchored boxes, *J. Complex. 14*, pp. 527–556.
- Morokofft William J., Cafisch Russel E. (1993). A quasi-Monte Carlo approach to particle Simulation of the heat equation. *SIAM, Society for Industrial and Applied Mathematics.*, pp. 1558–1573.
- Niederreitr, H. (1992), Random Number Generation and Quasi-Monte Carlo Methods, *CBMS-NSF Regional Conference Series in Applied Mathematics No. 63, SIAM, Philadelphia, PA*.
- Okten , G.,(1996), A probabilistic result on the discrepancy of a hybrid-Monte Carlo sequence and applications, *Monte Carlo Methods Appl. 2 (4)*, pp. 25–270.
- Okten, G.,(2001). High dimensional simulation, *Mathematics and Computers in Simulation 55* , pp. 215–222.
- Okten, G., Tuffin,B., Burago,V. (2006). A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance, *Journal of Complexity 22*, pp.435–458.
- Zdravko I. Botev, Taimre, T., Kroese, D.P. (2011). Handbook of Monte Carlo Methods. *Wiley, A John Wiley and Sons, Inc., Publication*.



On lower bounds for the variance of functions of random variables with weighted distributions

F. Goodarzi¹, M. Amini and G. R. Mohtashami Borzadaran
Ferdowsi University of Mashhad

Abstract: In this paper, based on the obtained characterization via the lower bounds for the variance of a function of random variable X , we find a characterization of the weighted function corresponding to density function $f(x)$, in terms of Chernoff-type inequalities. Subsequently, by using the new characterization, we derive the lower bounds for the variance of a function of the weighted random variable. Moreover, assuming that X is DFR [decreasing failure rate] we find an lower bound for the variance of this function with specific weight functions.

Keywords Variance bounds, Chernoff inequality, Size-biased distribution, decreasing failure rate.

Mathematics Subject Classification (2010): 60E15.

1 Introduction

During the last three decade, several papers have prepared regarding upper bounds for functions of random variables, since the appearance of the [Chernoff \(1981\)](#) inequality, giving an upper bound for the special case of a standard normal random variable. For any absolutely continuous function g ,

$$\text{Var}[g(X)] \leq E[g'(X)]^2. \quad (1.1)$$

[Chen \(1982\)](#) proved (1.1) using the integral representation of g and the Cauchy-Schwarz inequality. The equality in (1.1) holds if and only if (iff) the function g is a linear function of X . [Cacoullos \(1982\)](#) and [Cacoullos and Papathanasiou \(1989\)](#) obtained lower bounds for arbitrary random variables. [Cacoullos and Papathanasiou \(1995\)](#) and [Mohtashami and Shanbhag \(1998\)](#) established that, if X

¹Faranak Goodarzi: f-goodarzi@kashanu.ac.ir

is a continuous random variable with support an interval (a, b) , $-\infty \leq a < b \leq \infty$, mean μ , finite variance σ^2 and density function f , then the following general covariance identity holds:

$$\text{Cov}(h(X), g(X)) = E(z(X)g'(X)), \quad (1.2)$$

where g is an absolutely continuous function with $E|z(X)g'(X)| < \infty$, $h(x)$ is a given function and

$$z(x) = \frac{1}{f(x)} \int_a^x (E[h(X)] - h(t))f(t)dt. \quad (1.3)$$

(for $g(x) = x$, (1.2) yields $E[z(X)] = \text{Cov}(h(X), X)$).

They also established the converse of (1.2) and (1.3) reformulated as follows:

If there are functions $h(x)$ and $z(x)$ such that (1.2) holds for every differentiable g , then $h(x)$, $z(x)$ and the density f are related through (1.3).

If $h(x) = x$, then $z(x) = \sigma^2 w(x)$ and (1.2) reduces to the simple covariance identity [see [Cacoullos and Papathanasiou \(1989\)](#)]

$$\text{Cov}(X, g(X)) = \sigma^2 E[w(X)g'(X)], \quad (1.4)$$

where the w -function is defined by

$$\sigma^2 w(x)f(x) = \int_{-\infty}^x (\mu - t)f(t)dt. \quad (1.5)$$

[Cacoullos and Papathanasiou \(1997\)](#) showed that, under the conditions of identity (1.2), for every absolutely continuous function $h(x)$ with $h'(x) > 0$, the following inequality:

$$\text{Var}[g(X)] \geq \frac{E^2[z(X)g'(X)]}{E[z(X)h'(X)]}, \quad (1.6)$$

holds with equality iff $g(x) = c_1 h(x) + c_2$, where

$$z(x)f(x) = \int_{-\infty}^x (E[h(X)] - h(t))f(t)dt.$$

They, also established that if (a) there are functions h and z such that inequality (1.6) holds for every differentiable g and (b) equality $g(x) = c_1 h(x) + c_2$ holds, then h , z and f are related through (1.3).

Consider now a random variable X with density function $f(x)$ and distribution function $F(x)$. Let $\delta(x)$ be a non-negative function with finite non-zero expectation. Define a random variable X^* with

density function

$$f_{X^*}(x) = \frac{\delta(x)f(x)}{E[\delta(X)]}, \quad x \in \mathfrak{R}. \quad (1.7)$$

X^* is called the weighted random variable corresponding to X and its distribution is called the weighted distribution corresponding to $f(x)$. Though the concept of weighted distribution is due to Fisher (1934), it is Rao (1965) who studied the weighted distributions in a unified way.

In particular, if F is defined on $[0, \infty)$, then the weighted distribution with weight function $\delta(x) = x$ is called the size-biased or length-biased distribution. Indeed, since the distribution of X^* is weighted by the value, or size, of X , we say that X^* has the X -size biased distribution.

2 Main Results

In this paper, we give an lower bound for the variance of $g(X^*)$ by using characteristics of the associated random variable X . For this, we first obtain $w(\cdot)$ -function of random variable X^* , (i.e. $w^*(\cdot)$), according to $w(\cdot)$ -function of random variable X .

Theorem 2.1. *Let X be a continuous random variable with density function $f(x)$ and X^* be a weighted random variable with the weighted probability density function given by (1.7) and variance σ^{*2} . Then $w^*(\cdot)$ -function of random variable X^* , satisfies the equation*

$$\sigma^{*2}w^*(x)\delta(x) = \left\{ z_2(x) - z_1(x) \frac{E[X\delta(X)]}{E[\delta(X)]} \right\}, \quad (2.1)$$

where

$$z_1(x) = \frac{1}{f(x)} \int_{-\infty}^x (E[\delta(X)] - \delta(t))f(t)dt, \quad (2.2)$$

and

$$z_2(x) = \frac{1}{f(x)} \int_{-\infty}^x (E[X\delta(X)] - t\delta(t))f(t)dt. \quad (2.3)$$

Proof. Using (1.5), we have

$$\begin{aligned} \sigma^{*2}w^*(x)f_{X^*}(x) &= \int_{-\infty}^x \left(\frac{E[X\delta(X)]}{E[\delta(X)]} - t \right) \frac{\delta(t)}{E[\delta(X)]} f(t)dt \\ &= \frac{E[X\delta(X)]}{E^2[\delta(X)]} \int_{-\infty}^x \delta(t)f(t)dt - \int_{-\infty}^x \frac{t\delta(t)}{E[\delta(X)]} f(t)dt. \end{aligned}$$

Now, assuming that

$$z_1(x)f(x) = \int_{-\infty}^x (E[\delta(X)] - \delta(t))f(t)dt,$$

we get

$$\int_{-\infty}^x \delta(t)f(t)dt = E[\delta(X)]F_X(x) - z_1(x)f(x),$$

and by substitution, we conclude that

$$\begin{aligned} \sigma^{*2}w^*(x)f_{X^*}(x) &= \frac{1}{E[\delta(X)]} \int_{-\infty}^x (E[X\delta(X)] - t\delta(t))f(t)dt \\ &\quad - \frac{E[X\delta(X)]}{E^2[\delta(X)]} z_1(x)f(x) \\ &= \frac{z_2(x)f(x)}{E[\delta(X)]} - \frac{E[X\delta(X)]}{E^2[\delta(X)]} z_1(x)f(x), \end{aligned}$$

where

$$z_2(x)f(x) = \int_{-\infty}^x (E[X\delta(X)] - t\delta(t))f(t)dt,$$

consequently

$$\sigma^{*2}w^*(x)f_{X^*}(x) = \frac{f(x)}{E[\delta(X)]} \left\{ z_2(x) - z_1(x) \frac{E[X\delta(X)]}{E[\delta(X)]} \right\}, \tag{2.4}$$

This completes the proof of the theorem. □

In the special case, that X^* has the X -size biased distribution, we have

$$x\sigma^{*2}w^*(x) = \frac{1}{f(x)} \int_{-\infty}^x (E(X^2) - t^2)f(t) dt - \frac{E(X^2)}{E(X)} w(x)\sigma^2,$$

and thus

$$f(x) \left(x\sigma^{*2}w^*(x) + \frac{E(X^2)}{E(X)} w(x)\sigma^2 \right) = E(X^2)F(x) - \int_{-\infty}^x t^2 f(t)dt. \tag{2.5}$$

Remark 2.2. *If X is a continuous random variable and X^* has the weighted distribution given in (1.7), then a lower bound for the variance of a function g of the random variable X^* can be obtained by the following inequality*

$$\text{Var}[g(X^*)] \geq \frac{1}{\sigma^{*2}} \left(\frac{1}{E[\delta(X)]} E\{z_2(X)g'(X)\} - \frac{E[X\delta(X)]}{E^2[\delta(X)]} E\{z_1(X)g'(X)\} \right)^2, \tag{2.6}$$

where equality holds iff g is linear.

If X is a non-negative, integer-valued random variable with probability mass function $P(X = j)$ and X^* has the weighted distribution, given by

$$P(X^* = j) = \frac{\delta(j)}{E[\delta(X)]} P(X = j), \quad j = 0, 1, \dots,$$

then replacing integrals by sums and $g'(x)$ by $\Delta g(x) = g(x+1) - g(x)$, we arrive at the discrete version of (??) and (??) and the discrete version of (??) is denoted as follows:

$$\text{Var}[g(X^*)] \geq \frac{1}{\sigma^{*2}} \left(\frac{1}{E[\delta(X)]} E\{z_2(X)\Delta g(X)\} - \frac{E[X\delta(X)]}{E^2[\delta(X)]} E\{z_1(X)\Delta g(X)\} \right)^2, \quad (2.7)$$

where equality holds iff g is linear.

In the following proposition, we give a weighted distribution in which, probability density function has the general form $f(\cdot)$.

Proposition 2.3. *Let X_1, \dots, X_n be a random sample of size n from a continuous distribution with distribution function F and Y_r denote the r th-order statistic of this random sample. Then*

$$\text{Var}[g(Y_r)] \geq \frac{1}{\sigma^{*2}} \left(\frac{n!}{(r-1)!(n-r)!} E\{z_2(X)g'(X)\} - \left[\frac{n!}{(r-1)!(n-r)!} \right]^2 E[X\delta(X)] E\{z_1(X)g'(X)\} \right)^2, \quad (2.8)$$

where

$$z_1(x) = \frac{(r-1)!(n-r)!}{n!} \frac{1}{f(x)} \left\{ F(x) - \sum_{k=r}^n \binom{n}{k} [F(x)]^k [1-F(x)]^{n-k} \right\}, \quad (2.9)$$

and

$$\begin{aligned} z_2(x) = & \frac{1}{f(x)} \left[F(x) E\{X[F(X)]^{r-1} [1-F(X)]^{n-r}\} \right. \\ & - \frac{(r-1)!(n-r)!}{n!} \left\{ x \sum_{k=r}^n nk [F(x)]^k [1-F(x)]^{n-k} \right. \\ & \left. \left. - \sum_{k=r}^n nk \int_{-\infty}^x [F(t)]^k [1-F(t)]^{n-k} dt \right\} \right]. \end{aligned} \quad (2.10)$$

Proof. Obviously, the probability density function of Y_r ,

$$f_{Y_r}(x) = \frac{n!}{(r-1)!(n-r)!} [F(x)]^{r-1} [1-F(x)]^{n-r} f(x), \quad a < x < b,$$

belong to the family of weighted distributions with weight function $\delta(x) = [F(x)]^{r-1}[1 - F(x)]^{n-r}$ and $E[\delta(X)] = \frac{(r-1)!(n-r)!}{n!}$. Initially, we compute $z_1(x)f(x)$ and $z_2(x)f(x)$.

$$\begin{aligned} z_1(x)f(x) &= \frac{(r-1)!(n-r)!}{n!}F(x) - \int_{-\infty}^x [F(t)]^{r-1}[1 - F(t)]^{n-r}f(t)dt \\ &= \frac{(r-1)!(n-r)!}{n!}F(x) - \int_0^{F(x)} u^{r-1}(1-u)^{n-r}du \\ &= \frac{(r-1)!(n-r)!}{n!} \left\{ F(x) - \sum_{k=r}^n \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k} \right\}, \end{aligned}$$

and

$$\begin{aligned} z_2(x)f(x) &= F(x)E\left\{X[F(X)]^{r-1}[1 - F(X)]^{n-r}\right\} \\ &\quad - \int_{-\infty}^x t[F(t)]^{r-1}[1 - F(t)]^{n-r}f(t)dt, \end{aligned}$$

therefore, using the integration by parts, we obtain

$$\begin{aligned} z_2(x)f(x) &= F(x)E\left\{X[F(X)]^{r-1}[1 - F(X)]^{n-r}\right\} \\ &\quad - \frac{(r-1)!(n-r)!}{n!} \left\{ t \sum_{k=r}^n \binom{n}{k} [F(t)]^k [1 - F(t)]^{n-k} \Big|_{-\infty}^x \right\} \\ &\quad + \frac{(r-1)!(n-r)!}{n!} \sum_{k=r}^n \binom{n}{k} \int_{-\infty}^x [F(t)]^k [1 - F(t)]^{n-k} dt \\ &= F(x)E\left\{X[F(X)]^{r-1}[1 - F(X)]^{n-r}\right\} \\ &\quad - \frac{(r-1)!(n-r)!}{n!} \left\{ x \sum_{k=r}^n \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k} \right. \\ &\quad \left. - \sum_{k=r}^n \binom{n}{k} \int_{-\infty}^x [F(t)]^k [1 - F(t)]^{n-k} dt \right\}, \end{aligned}$$

if $\lim_{t \rightarrow -\infty} t \sum_{k=r}^n \binom{n}{k} [F(t)]^k [1 - F(t)]^{n-k} = 0$, which is equivalent to the restriction $\lim_{t \rightarrow -\infty} \tilde{r}(x) \neq 0$.

Eventually, substituting (2.9) and (2.10) into (2.6), the lower bound for the variance of $g(X^*)$ is obtained. \square

Remark 2.4. In proposition (2.3), Let X_1, \dots, X_n be a random sample from uniform distribution on $(0, 1)$ and $r = n$ then

$$z_1(x) = \frac{1}{n}x(1 - x^{n-1}), \quad z_2(x) = \frac{1}{n+1}x(1 - x^n).$$

Thus, the lower bound for variance of $g(Y_n)$ is calculated as follows:

$$\begin{aligned} \text{Var}[g(Y_n)] &\geq \frac{(n+1)^2(n+2)}{n} \left\{ n \int_0^1 g'(x) \frac{1}{n+1} x(1-x^n) dx \right. \\ &\quad \left. - \frac{n^2}{n+1} \int_0^1 g'(x) \frac{1}{n} x(1-x^{n-1}) dx \right\}^2 \\ &= n(n+2) \left\{ E[X^n(1-X)g'(X)] \right\}^2, \end{aligned}$$

and in the general case, the lower bound for the variance of $g(Y_r)$ will be

$$\text{Var}[g(Y_r)] \geq \frac{r(n-r+1)}{(n+1)^2(n+2)} \left\{ E \left[\frac{(n+2)!}{r!(n-r+1)!} X^r (1-X)^{n-r+1} g'(X) \right] \right\}^2.$$

At the end of this section, let F is DFR [decreasing failure rate], so that F is said to be DFR if $\bar{F}(x|t) = \frac{\bar{F}(x+t)}{\bar{F}(t)}$ is increasing in $0 \leq t < \infty$ for each $x \geq 0$. For more details see [Shaked and Shanthikumar \(2007\)](#).

Proposition 2.5. Let X be a non-negative random variable with density function $f(x)$, survival function $\bar{F}(x) = 1 - F(x)$ and failure (hazard) rate $r(x) = \frac{f(x)}{\bar{F}(x)}$. If F is DFR and X^* has a weighted distribution with weight function

$$(a) \delta(x) = \frac{f(x+t)}{f(x)}, \text{ then}$$

$$\text{Var}[g(X^*)] \geq \frac{1}{\sigma^{*2}} \left\{ E \left[\frac{X}{r(X)} g'(X) \right] \right\}^2, \quad (2.11)$$

$$(b) \delta(x) = I(x > t), \text{ then}$$

$$\text{Var}[g(X^*)] \geq \frac{1}{\sigma^{*2}} \left\{ E \left[\frac{X-t}{r(X)} g'(X) | X > t \right] \right\}^2, \quad (2.12)$$

such that in (2.11) and (2.12), $\sigma^{*2} = \text{Var}(X-t|X > t) = \frac{2}{\bar{F}(t)} \int_t^\infty \bar{F}(x)\mu(x)dx - \mu^2(t)$ is the variance residual life with $\mu(t) = E[X-t|X > t]$.

Proof. First we prove part (a). We know that, X^* is the residual life random variable at age t , denoted by $X_t = [X - t | X > t]$, and $E[X^*] = E[X - t | X > t] = \mu(t)$ and $\sigma^{*2} = Var[X - t | X > t]$. Now, by applying (1.5) and this point that $E[X - \mu] = 0$, we have

$$\begin{aligned} & \int_x^\infty (y - \mu(t)) \frac{\delta(y)}{E[\delta(X)]} f(y) dy = \int_x^\infty (y - \mu(t)) \frac{f(y+t)}{\bar{F}(t)} dy \\ &= \frac{1}{\bar{F}(t)} \left\{ \int_x^\infty y f(y+t) dy - \mu(t) \int_x^\infty f(y+t) dy \right\} \\ &= \frac{1}{\bar{F}(t)} \left\{ \int_{x+t}^\infty (y-t) f(y) dy - \mu(t) \int_{x+t}^\infty f(y) dy \right\} \\ &= \bar{F}(x|t) \left\{ \mu(x+t) - \mu(t) + x \right\}. \end{aligned}$$

Now, whereas if F is DFR then F is NWU (new worse than used), that is $\bar{F}(x|t) \geq \bar{F}(x)$ for each $x, t \geq 0$, and also F is IMRL, that is $\mu(s) \leq \mu(t)$ for $0 \leq s \leq t$, we have $\mu(x+t) \geq \mu(t)$ and consequently

$$w^*(x) \sigma^{*2} \frac{f(x+t)}{E[\delta(X)]} = \int_x^\infty (y - \mu(t)) \frac{\delta(y)}{E[\delta(X)]} f(y) dy \geq x \bar{F}(x),$$

which eventually leads to

$$\begin{aligned} Var[g(X^*)] &\geq \frac{1}{\sigma^{*2}} \left(\int_0^\infty x \bar{F}(x) g'(x) dx \right)^2 = \frac{1}{\sigma^{*2}} \left(\int_0^\infty \frac{x}{r(x)} g'(x) f(x) dx \right)^2 \\ &= \frac{1}{\sigma^{*2}} \left\{ E \left[\frac{X}{r(X)} g'(X) \right] \right\}^2, \end{aligned}$$

Now we prove part (b). It is obvious that $f_{X^*}(t) = \frac{I(x > t)}{P(X > t)} f(x)$ and so $E[X^*] = \mu(t) + t$ and $\sigma^{*2} = Var[X | X > t] = Var[X - t | X > t]$. Again by using (1.5), we have

$$\begin{aligned} \int_t^x (\mu(t) + t - y) \frac{f(t)}{\bar{F}(t)} dy &= \frac{\mu(t) + t}{\bar{F}(t)} \left\{ F(x) - F(t) \right\} \\ &\quad - \frac{1}{\bar{F}(t)} \left\{ \bar{F}(t)(\mu(t) + t) - \bar{F}(x)(\mu(x) + x) \right\} \\ &= \frac{\bar{F}(x)}{\bar{F}(t)} \left\{ \mu(x) - \mu(t) + (x - t) \right\}, \end{aligned}$$

and since F is IMRL, then

$$\int_t^x (\mu(t) + t - y) \frac{f(t)}{\bar{F}(t)} dy \geq \frac{\bar{F}(x)}{\bar{F}(t)} (x - t)$$

and therefore

$$\text{Var}[g(X^*)] \geq \frac{1}{\sigma^{*2}} \left(\int_t^\infty \frac{\bar{F}(x)}{\bar{F}(t)} (x-t) g'(x) dx \right)^2 = \frac{1}{\sigma^{*2}} \left\{ E \left[\frac{X-t}{r(X)} g'(X) | X > t \right] \right\}^2.$$

□

References

- Cacoullos, T. (1982). On upper and lower bounds for the variance of the function of a random variable, *The Annals of Probability*, **10**, 799–809.
- Cacoullos, T. and Papathanasiou, V. (1989). Characterizations of distributions by variance bounds, *Statistics and Probability Letters*, **7**, 351–356.
- Cacoullos, T. and Papathanasiou, V. (1995). A generalization of covariance identity and related characterization, *Mathematical Methods of Statistics* **4**, 106–113.
- Cacoullos, T. and Papathanasiou, V. (1997). Characterizations of distributions by generalizations of variance bounds and simple proofs of the CLT, *Journal of Statistical Planning and Inference*, **63**, 157–171.
- Chen, L.H.Y. (1982). An inequality for the multivariate normal distribution, *Journal of Multivariate Analysis*, **12**, 306–315.
- Chernoff H. (1981). A note on an inequality involving the normal distribution, *Annals of Probability*, **9**, 533–535.
- Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies, *The Annals of Eugenics*, **6**, 13–25.
- Mohtashami Borzadaran, G.R. and Shanbhag, D.N. (1998). Further results based on Chernoff-type inequalities, *Statistics and Probability Letters*, **39**, 109–117.

Rao, C.R. (1965). On discrete distributions arising out of methods of ascertainment. In: Patil, G.P. (Ed.), *Classical and Contagious Discrete Distributions* Permagon Press, Oxford and Statistical Publishing Society, Calcutta, pp. 320–332.

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders* (Springer, NewYork).



Usual Stochastic Order Between Largest Claim Amounts and its Applications in Insurance

Abedin Haidari¹, Ghobad Barmalzan and Amir T. Payandeh Najafabadi

Department of Mathematical Sciences, Shahid Beheshti University,

Abstract: Suppose $X_{\lambda_1}, \dots, X_{\lambda_n}$ is a set of non-negative random variables with X_{λ_i} having the distribution function $G(\lambda_i t)$, $\lambda_i > 0$ for $i = 1, \dots, n$, and I_{p_1}, \dots, I_{p_n} are independent Bernoulli random variables, independent of the X_{λ_i} 's, with $E(I_{p_i}) = p_i$, $i = 1, \dots, n$. Let $Y_i = I_{p_i} X_{\lambda_i}$, for $i = 1, \dots, n$. It is of interest to note that in actuarial science, it corresponds to the claim amount in a portfolio of risks. In this paper, we discuss the usual stochastic order between the largest claim amounts, $Y_{n:n}$, when the matrix of parameters $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda})$ changes to another matrix in a mathematical sense. We then apply the results for a special case of the scale model, namely, generalized gamma distribution with possibly different scale parameters to illustrate the established results.

Keywords: Usual Stochastic Order, Largest Claim Amount, Multivariate Chain Majorization, Scale Model, Generalized Gamma Distribution.

Mathematics Subject Classification (2010): 62N05.

1 Introduction

Largest claim amount is one of the main quantities in insurance companies. These largest claim amounts can change position and stability of them. In this regard, the largest claim amounts can have an important role in insurance analysis since they provide useful information for determining annual premium.

It is important for an actuary to be able to express preferences between random future gains or losses. For this purpose, stochastic orders have been used in various areas including management

¹Abedin Haidari: abedinhaidari@yahoo.com

science, financial economics, insurance, actuarial science, operations research, reliability theory, queuing theory and survival analysis. For example, in the decision theory of financial economics, we can assist an individual in making suitable decisions by comparing the risks leading to different uncertain payments. Interested readers may refer to [Shaked and Shanthikumar \(2007\)](#) for comprehensive discussions on univariate and multivariate stochastic orders.

The scale model is a flexible family of distributions that have been used extensively in statistics. Suppose G is an absolutely continuous distribution function with corresponding density function g . Then, independent random variables $X_{\lambda_1}, \dots, X_{\lambda_n}$ are said to belong to the scale model if $X_{\lambda_i} \sim G(\lambda_i t)$, where $\lambda_i > 0$ for $i = 1, \dots, n$. In this case, G and g are said to be the baseline distribution and density functions, respectively, and the parameters $\lambda_1, \dots, \lambda_n$ are referred to as the scale parameters. Assume that the baseline density function in the scale model has the form

$$g(t) = \frac{\beta}{\Gamma(\alpha)} t^{\alpha\beta-1} e^{-t^\beta}, \quad t > 0,$$

where $\alpha > 0$ and $\beta > 0$, and $\Gamma(\alpha)$ represents the complete gamma function. One may refer to [Stacy \(1962\)](#), [Johnson et. al \(1994\)](#) and [Balakrishnan and Peng \(2006\)](#) for some further details on this model.

The problem of comparison of number of claims and aggregate claim amounts with respect to some well-known stochastic orders is from interest on both theoretical and practical viewpoints. Several authors have worked on this problem; see, for example, [Karlin and Novikoff \(1963\)](#), [Ma \(2008\)](#), [Frostig \(2001\)](#), [Hu and Ruan \(2004\)](#), [Denuit and Frostig \(2006\)](#), [Khaledi and Ahmadi \(2008\)](#). Recently, [Barmalzan et al. \(2015\)](#) presented a complete version of the results of [Khaledi and Ahmadi \(2008\)](#) which have subsequently been extended to a more general case.

In this paper, we discuss the usual stochastic order between the largest claim amounts, $Y_{n:n}$, when the matrix of parameters $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda})$ changes to another matrix in a mathematical sense. We then apply the results for a special case of the scale model, namely, generalized gamma distribution

The rest of this paper is organized as follows. In Section 2, we introduce some definitions and notations pertinent to stochastic orders, multivariate majorization and vector majorization. Section 3 discuss stochastic comparisons of largest claim amounts with independent heterogeneous in scale models and then present some applications of these results.

2 Preliminaries

Definition 2.1. Suppose X and Y are two non-negative continuous random variables with distribution functions $F(t) = P(X \leq t)$ and $G(t) = P(Y \leq t)$, and survival functions $\bar{F}(t) = 1 - F(t)$ and $\bar{G}(t) = 1 - G(t)$, respectively. X is said to be larger than Y in the usual stochastic order (denoted by $X \geq_{st} Y$) if $\bar{F}(x) \geq \bar{G}(x)$ for all $x \in \mathbb{R}^+$.

Interested readers may refer to [Shaked and Shanthikumar \(2007\)](#) for comprehensive discussions on univariate and multivariate stochastic orders.

Definition 2.2. For two vectors $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$, let $\{a_{(1)}, \dots, a_{(n)}\}$ and $\{b_{(1)}, \dots, b_{(n)}\}$ denote the increasing arrangements of their components, respectively. Then, the vector \mathbf{a} is said to majorize the vector \mathbf{b} (denoted by $\mathbf{a} \succeq^m \mathbf{b}$) if $\sum_{j=1}^i a_{(j)} \leq \sum_{j=1}^i b_{(j)}$ for $i = 1, \dots, n-1$, and $\sum_{j=1}^n a_{(j)} = \sum_{j=1}^n b_{(j)}$.

Definition 2.3. A real-valued function ϕ , defined on a set $\mathbb{A} \subseteq \mathbb{R}^n$, is said to be Schur-convex (Schur-concave) on \mathbb{A} if $\mathbf{a} \succeq^m \mathbf{b}$ implies $\phi(\mathbf{a}) \geq (\leq) \phi(\mathbf{b})$ for any $\mathbf{a}, \mathbf{b} \in \mathbb{A}$.

A square matrix Π_n , of order n , is said to be a permutation matrix if each row and column has a single entry as 1, and all other entries as zero. Such matrices are obtained by interchanging rows (or columns) of the identity matrix I_n . The T-transform matrix has the form $T = wI_n + (1-w)\Pi_n$, where $0 \leq w \leq 1$, and Π_n is a permutation matrix that just interchanges two coordinates. It should be mentioned here that permutation matrices used in the T-transform matrices are the identity matrix I_n with its two columns interchanged. We say that two T-transform matrices $T_1 = w_1I_n + (1-w_1)\Pi_1$ and $T_2 = w_2I_n + (1-w_2)\Pi_2$ have the same structure if $\Pi_1 = \Pi_2$. It is well-known that the finite product of T-transform matrices with the same structures is also a T-transform matrix, while this product may not be a T-transform matrix if its elements do not have the same structure; see [Balakrishnan et al. \(2015\)](#) for more details.

The following definition presents two generalizations of the vector majorization.

Definition 2.4. Consider the $m \times n$ matrices $U = \{u_{ij}\}$ and $V = \{v_{ij}\}$ with rows $\mathbf{u}_1, \dots, \mathbf{u}_m$ and $\mathbf{v}_1, \dots, \mathbf{v}_m$, respectively. Then:

- (i) If $\mathbf{u}_i \stackrel{m}{\succeq} \mathbf{v}_i$ for $i = 1, \dots, m$, then we say that the matrix U is larger than the matrix V in row majorization, denoted by $U >^{row} V$;
- (ii) If there exists a finite set of $n \times n$ T -transform matrices T_1, \dots, T_k such that $V = UT_1T_2 \dots T_k$, then we say that the matrix U is larger than matrix V in the chain majorization, denoted by $U \gg V$.

It is well-known that chain majorization implies row majorization. Interested readers may refer to Marshall et al. (2011) for an elaborate discussion on the theory of vector and matrix majorizations and their applications.

Lemma 2.5. (Balakrishnan et al. (2015)) Consider the differentiable function $\phi : R^{+4} \rightarrow R^+$. Then,

$$\phi(U) \geq (\leq) \phi(V) \quad \text{for all } U, V \text{ such that } U \in \mathcal{S}_2, \text{ and } U \gg V \quad (2.1)$$

if and only if

- (i) $\phi(U) = \phi(U\Pi)$ for all permutation matrices Π , and for all $U \in \mathcal{U}_2$;
- (ii) $\sum_{i=1}^2 (u_{ik} - u_{ij})[\phi_{ik}(U) - \phi_{ij}(U)] \geq (\leq) 0$ for all $j, k = 1, 2$, and for all $U \in \mathcal{U}_2$, where $\phi_{ij}(U) = \partial\phi(U)/\partial u_{ij}$.

Lemma 2.6. (Balakrishnan et al. (2015)) Consider the differentiable function $\varphi : R^{+2} \rightarrow R^+$, and let the function $\phi_n : R^{+2n} \rightarrow R^+$ be defined as $\phi_n(U) = \prod_{i=1}^n \varphi(u_{1i}, u_{2i})$. If ϕ_2 satisfies (2.1), then $\phi_n(U) \geq \phi_n(V)$, where $U \in \mathcal{U}_n$ and $V = UT$.

3 Results for largest claim amounts

Note that the random variables $Y_i = I_{p_i} X_{\lambda_i}$, $i = 1, \dots, n$, are discrete-continuous type, which admit zero with probability $1 - p_i$, and X_{λ_i} with probability p_i , $i = 1, \dots, n$. Then, the distribution function of $Y_{n:n}$, the largest claim amount, is given by

$$F_{Y_{n:n}}(t) = \prod_{i=1}^n (1 - p_i \bar{G}(\lambda_i t)), \quad t \geq 0. \quad (3.1)$$

Hereafter, we assume that $Y_{n:n}^*$ denotes similarly the largest claim amount arising from $Y_i^* = I_{p_i^*} X_{\lambda_i^*}$, $i = 1, \dots, n$, where $X_{\lambda_1^*}, \dots, X_{\lambda_n^*}$ are independent non-negative random variables with $X_{\lambda_i^*} \sim G(\lambda_i^* t)$, $i = 1, \dots, n$, and $I_{p_1^*}, \dots, I_{p_n^*}$ is a set of independent random variables, independent of the $X_{\lambda_i^*}$'s, with $E(I_{p_i^*}) = p_i^*$, $i = 1, \dots, n$. So, the survival function of $Y_{n:n}^*$ can be obtained readily from (3.1) upon replacing p_i by p_i^* and λ_i by λ_i^* , $i = 1, \dots, n$.

3.1 Ordering results

Let us set

$$\mathcal{S}_n = \left\{ (\mathbf{a}, \mathbf{b}) = \begin{pmatrix} a_1, \dots, a_n \\ b_1, \dots, b_n \end{pmatrix} : a_i, b_j > 0 \text{ and } (a_i - a_j)(b_i - b_j) \geq 0, i, j = 1, \dots, n \right\}.$$

Theorem 3.1. *Suppose $X_{\lambda_1}, X_{\lambda_2}$ are independent non-negative random variables with $X_{\lambda_i} \sim G(\lambda_i t)$, $i = 1, \dots, n$, where G is an absolutely continuous distribution function with corresponding density function g . Further, suppose I_{p_1}, I_{p_2} is a set of independent Bernoulli random variables, independent of the X_{λ_i} 's, with $E(I_{p_i}) = p_i$, $i = 1, 2$. Assume that the following conditions hold:*

- (i) $g(t)$ is decreasing in t ;
- (ii) $h : [0, 1] \rightarrow (0, \infty)$ is differentiable and a strictly decreasing convex function.

Then, for $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda}) \in \mathcal{S}_2$, we have

$$\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \end{pmatrix} \implies Y_{2:2} \geq_{st} Y_{2:2}^*.$$

Proof. According to (??), the distribution function of $Y_{2:2}$ can be expressed as

$$F_{Y_{2:2}}(t) = \prod_{i=1}^2 \left(1 - h^{-1}(u_i) \bar{G}(\lambda_i t) \right), \quad t \geq 0,$$

where h^{-1} denotes the inverse of h , and $\mathbf{u} = (u_1, u_2)$ with $u_i = h(p_i)$, $i = 1, 2$. To establish the desired result, we have to check Conditions (i) and (ii) of Lemma 2.5. Clearly, $F_{Y_{2:2}}(t)$ is permutation invariant

on \mathcal{S}_2 , for fixed $t \geq 0$, which confirms Condition (i) of Lemma 2.5. On the other hand, for fixed $t \geq 0$ and $i \neq j$, consider the function φ as

$$\begin{aligned} \varphi(\mathbf{u}, \boldsymbol{\lambda}) &= (u_1 - u_2) \left(\frac{\partial F_{Y_{2:2}}(t)}{\partial u_1} - \frac{\partial F_{Y_{2:2}}(t)}{\partial u_2} \right) \\ &\quad + (\lambda_1 - \lambda_2) \left(\frac{\partial F_{Y_{2:2}}(t)}{\partial \lambda_1} - \frac{\partial F_{Y_{2:2}}(t)}{\partial \lambda_2} \right). \end{aligned} \quad (3.2)$$

The partial derivative of $F_{Y_{2:2}}(t)$ with respect to λ_i and u_i are

$$\frac{\partial F_{Y_{2:2}}(t)}{\partial u_i} = -F_{Y_{2:2}}(t) \frac{\bar{G}(\lambda_i t) \left(\frac{\partial h^{-1}(u_i)}{\partial u_i} \right)}{1 - h^{-1}(u_i) \bar{G}(\lambda_i t)} \quad \text{and} \quad \frac{\partial F_{Y_{2:2}}(t)}{\partial \lambda_i} = t F_{Y_{2:2}}(t) \frac{h^{-1}(u_i) g(\lambda_i t)}{1 - h^{-1}(u_i) \bar{G}(\lambda_i t)},$$

respectively. Now, upon substituting these derivatives in (3.2), we obtain

$$\begin{aligned} \varphi(\mathbf{u}, \boldsymbol{\lambda}) &\stackrel{sgn}{=} (u_1 - u_2) \left(\frac{\bar{G}(\lambda_2 t) \left(\frac{\partial h^{-1}(u_2)}{\partial u_2} \right)}{1 - h^{-1}(u_2) \bar{G}(\lambda_2 t)} - \frac{\bar{G}(\lambda_1 t) \left(\frac{\partial h^{-1}(u_1)}{\partial u_1} \right)}{1 - h^{-1}(u_1) \bar{G}(\lambda_1 t)} \right) \\ &\quad + t(\lambda_1 - \lambda_2) \left(\frac{h^{-1}(u_1) g(\lambda_1 t)}{1 - h^{-1}(u_1) \bar{G}(\lambda_1 t)} - \frac{h^{-1}(u_2) g(\lambda_2 t)}{1 - h^{-1}(u_2) \bar{G}(\lambda_2 t)} \right), \end{aligned} \quad (3.3)$$

where $a \stackrel{sgn}{=} b$ means that a and b have the same sign. The assumption that $(\mathbf{u}, \boldsymbol{\lambda}) \in \mathcal{S}_2$ implies that $(u_1 - u_2)(\lambda_1 - \lambda_2) \geq 0$, which means that $u_1 \geq u_2$ and $\lambda_1 \geq \lambda_2$, or $u_2 \geq u_1$ and $\lambda_2 \geq \lambda_1$. We present the proof only for the case when $u_1 \geq u_2$ and $\lambda_1 \geq \lambda_2$, since the proof for the other case is quite similar. Note that, since h is strictly decreasing and convex, h^{-1} is also strictly decreasing and convex.

We then have

$$\begin{aligned} \bar{G}(\lambda_2 t) \left(\frac{\partial h^{-1}(u_2)}{\partial u_2} \right) &\leq \bar{G}(\lambda_1 t) \left(\frac{\partial h^{-1}(u_2)}{\partial u_2} \right) \\ &\leq \bar{G}(\lambda_1 t) \left(\frac{\partial h^{-1}(u_1)}{\partial u_1} \right) < 0, \end{aligned} \quad (3.4)$$

where the first inequality follows from the facts that both functions \bar{G} and h^{-1} are decreasing, while the second inequality follows from the convexity of h^{-1} . Furthermore, from the decreasing property of both functions \bar{G} and h^{-1} once again, it can be easily concluded that

$$0 \leq h^{-1}(u_1) \bar{G}(\lambda_1 t) \leq h^{-1}(u_2) \bar{G}(\lambda_2 t) \leq 1. \quad (3.5)$$

Now, from (3.4) and (3.5), we readily observe that the first term on the right hand side of (3.3) is non-positive. Now, from assumption (i) and the facts that $u_1 \geq u_2$ and $\lambda_1 \geq \lambda_2$, it follows that

$$h^{-1}(u_1)g(\lambda_1 t) \leq h^{-1}(u_2)g(\lambda_2 t),$$

which together with (3.5) imply that the second term on the right hand side of (3.3) is also non-positive. Upon combining the obtained results, we have $\varphi(\mathbf{u}, \boldsymbol{\lambda}) \leq 0$, and so Condition (ii) of Lemma 2.5 is satisfied. This completes the proof of the theorem. \square

Some generalizations of the result in Theorem 3.1 to the case when the number of underlying random variables is arbitrary are presented below.

Theorem 3.2. *Suppose $X_{\lambda_1}, \dots, X_{\lambda_n}$ are independent non-negative random variables with $X_{\lambda_i} \sim G(\lambda_i t)$, $i = 1, \dots, n$, where G is an absolutely continuous distribution function with corresponding density function g . Further, suppose I_{p_1}, \dots, I_{p_n} is a set of independent Bernoulli random variables, independent of the X_{λ_i} 's, with $E(I_{p_i}) = p_i$, $i = 1, \dots, n$. Assume that the following conditions hold:*

(i) $g(t)$ is decreasing in t ;

(ii) $h : [0, 1] \rightarrow (0, \infty)$ is differentiable and a strictly decreasing convex function.

Then, for $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda}) \in \mathcal{S}_n$, we have

$$\begin{pmatrix} h(p_1^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \cdots & \lambda_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & \cdots & h(p_n) \\ \lambda_1 & \cdots & \lambda_n \end{pmatrix} T \implies Y_{n:n} \geq_{st} Y_{n:n}^*.$$

Proof. Setting $\phi_n(\mathbf{p}, \boldsymbol{\lambda}) = F_{Y_{n:n}}(t)$ and $\varphi(p, \lambda) = 1 - p\bar{G}(\lambda t)$, for fixed $t \geq 0$, we then have $\phi_n(\mathbf{p}, \boldsymbol{\lambda}) = \prod_{i=1}^n \varphi(p_i, \lambda_i)$. According to Theorem 3.1, ϕ_2 is satisfied in (2.1). Now, the required result follows immediately from Lemma 2.6. \square

Theorem 3.3. *Under the assumptions of Theorem 3.2, if the T -transform matrices T_1, \dots, T_k have the same structure, then for $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda}) \in \mathcal{S}_n$, then we have*

$$\begin{pmatrix} h(p_1^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \cdots & \lambda_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & \cdots & h(p_n) \\ \lambda_1 & \cdots & \lambda_n \end{pmatrix} T_1 \cdots T_k \implies Y_{n:n} \geq_{st} Y_{n:n}^*.$$

Proof. The desired result is immediately obtained from Theorem 3.2 and the fact that the finite product of T -transform matrices with the same structure is also a T -transform matrix. \square

It will be of interest to know whether the results of Theorem 3.3 will still hold if the matrices T_i , $i = 1, \dots, k$, do not have the same structure. The following theorem gives an answer to this question.

Theorem 3.4. *Under the assumptions of Theorem 3.2, if $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda}) \in \mathcal{S}_n$ and $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda})T_1 \cdots T_i \in \mathcal{S}_n$ for $i = 1, \dots, k - 1$, where $k \geq 2$, then we then have*

$$\begin{pmatrix} h(p_1^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \cdots & \lambda_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & \cdots & h(p_n) \\ \lambda_1 & \cdots & \lambda_n \end{pmatrix} T_1 \cdots T_k \implies Y_{n:n} \geq_{st} Y_{n:n}^*.$$

Proof. The required result can be easily obtained by repeating the result of Theorem 3.2 for the matrices $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda})T_1 \cdots T_i$, $i = 1, \dots, k - 1$. \square

3.2 Some Applications

Lemma 3.5. *Let $X \sim GG(\alpha, \beta, 1)$ with density function g . Then, for $\alpha\beta \leq 1$, $g(t)$ is decreasing in $t \in (0, \infty)$.*

Proof. Taking derivative of $g(t)$ with respect to t , it readily follows that

$$g'(t) \stackrel{sgn}{=} \alpha\beta - 1 - \beta t^\beta = \phi(t), \text{ say.}$$

Clearly, $\phi(t)$ is decreasing in $t \in (0, \infty)$ which results in $\phi(t) \leq \phi(0) = \alpha\beta - 1$ for all $t > 0$. Now, from the assumption $\alpha\beta \leq 1$, the desired result follows. \square

The following corollary, which is a direct consequence of Theorem 3.3 and Lemma 3.5, enables us to compare the largest claim amounts in the sense of the usual stochastic order.

Corollary 3.6. *Assume that $X_{\lambda_i} \sim GG(\alpha, \beta, \lambda_i)$, $i = 1, \dots, n$. Under the assumption of Theorem 3.3, if the T -transform matrices T_1, \dots, T_k have the same structure, then for $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda}) \in \mathcal{S}_n$, then for $\alpha\beta \leq 1$ we have*

$$\begin{pmatrix} h(p_1^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \cdots & \lambda_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & \cdots & h(p_n) \\ \lambda_1 & \cdots & \lambda_n \end{pmatrix} T_1 \cdots T_k \implies Y_{n:n} \geq_{st} Y_{n:n}^*.$$

The following corollary is immediately obtained from Theorem 3.4 and Lemma 3.5.

Corollary 3.7. *Under the assumption of Corollary 3.6, assume that $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda}) \in \mathcal{S}_n$ and $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda})T_1 \cdots T_i \in \mathcal{S}_n$ for $i = 1, \dots, k-1$, where $k \geq 2$. If $\alpha\beta \leq 1$ and $h : [0, 1] \rightarrow (0, \infty)$ is decreasing and convex, then for $(\mathbf{h}(\mathbf{p}), \boldsymbol{\lambda}) \in \mathcal{S}_n$, we have*

$$\begin{pmatrix} h(p_1^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \cdots & \lambda_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & \cdots & h(p_n) \\ \lambda_1 & \cdots & \lambda_n \end{pmatrix} T_1 \cdots T_k \implies Y_{n:n} \geq_{st} Y_{n:n}^*.$$

References

- Balakrishnan, N., Haidari, A. and Masoumifard, K. (2015). Stochastic comparisons of series and parallel systems with generalized exponential components. *IEEE Transactions on Reliability*, **64**, 333-348.
- Balakrishnan and Peng (2006). Generalize gamma model. *Statistics in Medicine*, **25**, 2797-2816.
- Barmalzan, G., Payandeh Najafabadi, A.T. and Balakrishnan, N. (2015). Stochastic comparison of aggregate claim amounts between two heterogenous portfolios and its applications, *Insurance: Mathematics and Economics*, **61**, 235-241.
- Denuit, M. and Frostig, E. (2006). Heterogeneity and the need for capital in the individual model, *Scandinavian Actuarial Journal*, **1**, 42-66.
- Frostig, E. (2001). A comparison between homogeneous and heterogeneous portfolios, *Insurance: Mathematics and Economics*, **29**, 59-71.
- Hu, T. and Ruan, L. (2004). A note on multivariate stochastic comparisons of Bernoulli random variables, *Journal of Statistical Planning and Inference*, **126**, 281-288.
- Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol 1*, Second edition. John Wiley & Sons, New York.
- Karlin, S. and Novikoff, A. (1963). Generalized convex inequalities, *Pacific Journal of Mathematics*, **13**, 1251-1279.

-
- Khaledi, B. and Ahmadi, S.S. (2008). On stochastic comparison between aggregate claim amounts, *Journal of Statistical Planning and Inference*, **138**, 3121-3129.
- Ma, C. (2000). Convex orders for linear combinations of random variables, *Journal of Statistical Planning and Inference*, **84**, 11-25.
- Marshall, A.W., Olkin, I. and Arnold, B.C. (2011). *Inequalities: Theory of Majorization and its Applications*, Second edition, Springer, New York.
- Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*, Springer, New York.
- Stacy, E.W. (1962). A generalization of the gamma distribution. *Annals of Mathematical Statistics*, **33**, 1187-1192.



Periodically correlated space-time autoregressive Hilbertian processes of order one

M. Hashemi⁴, J. Mateu²

¹Department of statistics, University of Isfahan, Khansar Unit, Isfahan, Iran

² Department of Mathematics, university JAUME

Abstract: In this article, We consider periodically correlated space-time autoregressive processes in Hilbert spaces. We present here results involving existence, and the strong law of large numbers.

Keywords periodically correlated space time autoregressive Hilbertian processes, periodically correlated autoregressive Hilbertian processes, T-periodic sequences, strong law of large number

Mathematics Subject Classification (2010): 60G20.

1 Introduction

Spatial time series is addressed to the data that depend on time and spatial location. It deals with usually single variable observed over time at a number of different locations. Such kinds of data are also recognized as space-time data.

The Space-Time Autoregressive Moving-Average (STARMA) model class developed by Pfeifer and Deutsch (1980a, 1980b, 1981a, 1981b, 1981c). Processes that can be modelled by the STARMA models are characterized by a single random variable observed at N fixed sites in space. The dependencies between the N time series is incorporated in the model through hierarchial $N \times N$ weighting matrices, specified by the model builder prior to analyzing the data. These weighting matrices should incorporate the relevant physical characteristics of the system into the model. Each of the N time series is simultaneously modelled as a linear combination of past observations and disturbances as well as

⁴M. Hashemi : maryam3752@yahoo.com

weighted past observations and disturbances at neighboring sites.

A STARMA(p_λ, q_m) model is specified as

$$y_t = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} W_l y_{t-k} - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} W_l \epsilon_{t-k} + \epsilon_t$$

where the p, q are the temporal AR and MA lags, while λ, m are the spacial lags, W_l is the $N \times N$ matrix of weights for spatial order l . ϵ_t is the random normally distributed disturbance at time t .

This model has been applied to numerous applications ranging from environmental (Pfeifer and Deutsch 1981a, Stofer 1986) to epidemiological (Pfeifer and Deutsch 1980a) and economic (Pfeifer and Bodily 1990).

Now a day, in many areas it is common to work with large databases, which often increase these observations of a random variable taken over a continuous interval (or in increasingly larger discretizations of a continuous interval). Functional data analysis has emerged as a significant tool for modeling large dimension data in the last decade. From the practical standpoint it started as a method of utilizing large data sets that have become available characterized by a large record frequency and a limited number of periods when the data has been obtained.

The Hilbertian autoregressive model of order 1 (ARH(1)) generalizes the classical AR(1) model to random elements with values in Hilbert spaces. This model was introduced by Bosq (2000), then studied by several authors, as Mourid (1993), Besse and Cardot (1996), Pumo (1999), Mas (2002, 2007), Horvath et al. (2010). Bosq in his fundamental work (2000) provides basic results on Hilbertian strongly second order autoregressive and moving average processes. The existence, covariance structure, parameter estimation, strong law of large numbers and central limit theorem, are the topics that are covered by Bosq (2000).

High dimensional statistics has been developed in context of spatial statistics too (Bosq, 2000; Ruiz-Medina and Salmeron 2010).

In the spatial functional time series context, Ruiz-Medina (2012) introduced and studied the structural properties of a class of spatial autoregressive Hilbertian processes, that called SARH(1) processes. The spatial moving average Hilbertian representation is also studied. The formulation of a SARH(1) plugin

extrapolator is achieved in Ruiz-Medina (2012) from the method of moments.

The periodically correlated autoregressive Hilbertian process of order one (PCARH(1)) was introduced by Hashemi and Soltani (2011). They first furnished the structure and existence: imbedding every periodically correlated autoregressive Hilbertian process in the class of autoregressive Hilbertian processes of higher dimensions, providing necessary and sufficient conditions for the existence of these processes. The topics that they presented for periodically correlated autoregressive Hilbertian processes of order one include: law of large numbers; Central limit Theorem; The variance and covariance estimations; based on given segments of these processes.

In this article we introduce periodically correlated space-time autoregressive Hilbertian processes and provide their existence and strong law of large numbers. Our methodology is to imbed every PCSTARH(1) of period T in \mathcal{H} into a subclass of PCARH(1) processes then benefit from Hashemi and Soltani (2011) derivations. This article is organized as follows. We introduce PCSTARH(1) processes and provide their existence in section 2. We give the strong large number in section 3.

2 PCSTARH(1) Processes

For writing definition, example, lemma, theorem, proof, Throughout this paper, we consider \mathcal{H} as a real separable Hilbert space equipped with scalar product $\langle \cdot, \cdot \rangle$, norm $\| \cdot \|$ and Borel σ -field \mathcal{B} . The \mathcal{H} -valued random variables considered below are defined over the same probability space (Ω, \mathcal{F}, P) supposed to be rich enough and complete.

We first define periodically correlated Hilbertian white noise and a T -periodic bounded linear operator sequence.

Definition 2.1. *A periodically correlated H -valued process $\{\epsilon_{it}, i = 1, \dots, k, t \in \mathbb{Z}\}$ is white noise if it satisfies the following properties*

- i) $E\epsilon_{it} = 0$, $0 < E \| \epsilon_{it} \|^2 = \sigma_{it}^2 < \infty$ for every $i = 1, \dots, k, t \in \mathbb{Z}$*
- ii) $C_{\epsilon_{it}} = C_{\epsilon_{i(t+T)}}$ for every $i = 1, \dots, k, t \in \mathbb{Z}$*
- iii) $C_{\epsilon_{it}, \epsilon_{jl}} = 0$ for all $t \neq l$ or $i \neq j$*

Definition 2.2. A sequence $\{\rho_t, t \in \mathbb{Z}\}$ in $\mathcal{L}(\mathcal{H})$ is called T -periodic if $\rho_t = \rho_{t+T}$.

Definition 2.3. A Hilbertian process $\{X_{it}, i = 1, \dots, k, t \in \mathbb{Z}\}$ is called periodically correlated space-time autoregressive Hilbertian process of order one (PCSTARH(1,1)) with period T , if it is periodically correlated and satisfies

$$X_{it} = \phi_t X_{i(t-1)} + \psi_t \sum_{j=1}^k w_{ij} X_{j(t-1)} + \epsilon_{it} \tag{2.1}$$

Where $\{\phi_t, t \in \mathbb{Z}\}, \{\psi_t, t \in \mathbb{Z}\}$ are T -periodic sequences in $\mathcal{L}(\mathcal{H})$ and $\{w_{ij}, i = 1, \dots, k, j = 1, \dots, k\}$ are bounded linear operators in \mathcal{H} and $\{\epsilon_{it}, i = 1, \dots, k, t \in \mathbb{Z}\}$ is a PCHWN.

Note that we define \mathbf{X}_t as a \mathcal{H}^k -valued random variable such that $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{kt})'$ and $W = (w_{ij})$ is a $k \times k$ matrix in which each element is a bounded linear operator and $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{kt})'$. Now we can rewrite our model as below

$$\mathbf{X}_t = \phi_t \mathbf{X}_{(t-1)} + \psi_t W X_{(t-1)} + \epsilon_t \tag{2.2}$$

Suppose that I stands for the identity operator on \mathcal{H}^k , $I\mathbf{x} = \mathbf{x}$ for each x in \mathcal{H}^k . So we can rewrite our model as below

$$\mathbf{X}_t = (\phi_t I + \psi_t W)\mathbf{X}_{(t-1)} + \epsilon_t \tag{2.3}$$

Corollary 2.4. The Hilbertian process $\mathbf{X}_t = (\phi_t I + \psi_t W)\mathbf{X}_{(t-1)} + \epsilon_t$, is a PCARH(1) process where $\{\phi_t, t \in \mathbb{Z}\}, \{\psi_t, t \in \mathbb{Z}\}$ are T -periodic sequences in $\mathcal{L}(\mathcal{H})$ and W is a bounded linear operator in \mathcal{H}^k .

Proof. We first show that $\rho_t = \phi_t I + \psi_t W$ is a T -periodic sequence in $\mathcal{L}(\mathcal{H}^k)$. Since ϕ_t, ψ_t are T -periodic bounded linear operators, we have

$$\rho_{t+T} = \phi_{t+T} I + \psi_{t+T} W = \phi_t I + \psi_t W$$

It is clear that ϵ_t is PCHWN, because

- 1) $E\epsilon_t = E(\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{kt})' = 0,$
- 2) $C_{\epsilon_{n+T}}(x) = E\langle \epsilon_{n+T}, x \rangle \epsilon_{n+T} = E(\sum_{i=1}^k \langle \epsilon_{i(n+T)}, x_i \rangle \epsilon_{1(n+T)}, \sum_{i=1}^k \langle \epsilon_{i(n+T)}, x_i \rangle \epsilon_{2(n+T)}, \dots, \sum_{i=1}^k \langle \epsilon_{i(n+T)}, x_i \rangle \epsilon_{k(n+T)})' = E(\sum_{i=1}^k \langle \epsilon_{in}, x_i \rangle \epsilon_{1n}, \sum_{i=1}^k \langle \epsilon_{in}, x_i \rangle \epsilon_{2n}, \dots, \sum_{i=1}^k \langle \epsilon_{in}, x_i \rangle \epsilon_{kn})' = C_{\epsilon_n}(x)$

3) $C_{\epsilon_n, \epsilon_m} = 0$ for each $n \neq m$.

The proof is complete. \square

Assumption \mathbf{A}_1 : There are integers $k_0, k_1, \dots, k_{T-1} \in [1, \infty)$ such that $\sum_{i=0}^{T-1} \|\rho_i\|^{k_i} < 1$, where $\rho_i = \phi_i I + \psi_i W$.

Corollary 2.5. *If $\sum_{i=0}^{T-1} 2^{k_i} (\|\rho_i\|^{k_i} + \|\psi_i\|^{k_i} \|W\|^{k_i}) < 1$, then assumption \mathbf{A}_1 holds.*

Proof. It is enough that we use the known result

$$(x + y)^p \leq 2^p (x^p + y^p)$$

Where x, y and p are positive. \square

We may now give a statement concerning existence and uniqueness of \mathbf{X} .

Theorem 2.6. *Under the assumption \mathbf{A}_1 , the equation $\mathbf{X}_t = (\phi_t I + \psi_t W)\mathbf{X}_{t-1} + \epsilon_t$ has a unique solution given by*

$$\mathbf{X}_{nT+i} = \sum_{j=0}^{\infty} A_{j,nT+i} \epsilon_{nT+i-j} = \sum_{k=0}^{\infty} \sum_{l=0}^{T-1} [A_{T,nT+i}]^k A_{l,nT+i} \epsilon_{(n-k)T+i-l}, \quad (2.4)$$

Where $A_{0,i} = I, A_{1,i} = \rho_i, A_{2,i} = \rho_i \rho_{i-1}, \dots, A_{k,i} = \rho_i \rho_{i-1} \dots \rho_{i-k+1}$ and $\rho_i = \phi_i I + \psi_i W$.

Proof. Based on Corollary 2.1, \mathbf{X}_t is a PCARH(1) process, by using Theorem 2.1 of Soltani and Hashemi (2011), consider $\rho_t = \phi_t I + \psi_t W$ so we have $\mathbf{X}_t = \rho_t \mathbf{X}_{t-1} + \epsilon_t$. Under the assumption \mathbf{A}_1 a PCARH(1) processes have a unique solution as equation 4 and the proof is complete. \square

3 Strong law of large number

In this section, we prove strong law large number for PCARH(1) processes.

Definition 3.1. *A PCSTARH(1,1), $\{X_{it}, i = 1, \dots, k, t \in \mathbb{Z}\}$, is said to be standard if assumption \mathbf{A}_1 holds.*

Theorem 3.2. Let $\{X_{it}, i = 1, \dots, k, t \in \mathbb{Z}\}$ be a standard PCSTARH(1,1) and $X_{i,0}, X_{i,1}, \dots, X_{i,NT-1}$ be a finite segment of this model. Then as $N \rightarrow \infty$,

$$\frac{n^{\frac{1}{4}}}{(\log n)^{\beta}} \frac{S_{i,n}(X)}{n} \xrightarrow{a.s.} 0, \quad \beta > \frac{1}{2}$$

Where $S_{i,n}(X) = \sum_{t=0}^{n-1} X_{it}$ and $n = NT$.

Proof. By defining $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{kt})'$, X_t is a PCARH(1) process and by using Theorem (3.1) of Soltani and Hashemi (2011), we have

$$\frac{n^{\frac{1}{4}}}{(\log n)^{\beta}} \frac{S_n(X)}{n} \xrightarrow{a.s.} 0, \quad \beta > \frac{1}{2}.$$

and the proof is complete. □

References

- Besse, P. and Cardot H. (1996). "Approximation spline de la prevision d'un processus fonctionnel autoregressif d'ordre." *Canad. J. Statist.*, Vol. 24, 467-487.
- Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Vol. 149, Berlin, Springer.
- Hagbin, H., Shishebor, Z. and Soltani, A. R. (2014). "Hilbertian spatial periodically correlated first order autoregressive models." *Advances in Data Analysis and Classification*, Vol. 8(3), 303- 319.
- Horvath, L., Huskova, M. and Kokoszka, P. (2010). "Testing the stability of the functional autoregressive process." *J. Multivariate Anal.*, Vol. 101, 352-367.
- Mas, A. (2002). "Weak convergence for the covariance operators of a Hilbertian linear process." *Stochastic Process Appl.*, Vol. 99, 117-135.
- Mas, A. (2007). "Weak convergence in the functional autoregressive model." *J. Multivariate Anal.*, Vol. 98, 1231-1261.
- Mourid, T. (1993). "Processus autoregressifs d'ordre superieur." *Acad. Sci.*, t.317, Ser. I, 167-1172.

- Pfeifer, P. Deutsch S. (1980a). "A three stage iterative procedure for space time modeling." *Technometrics*, Vol. 22, No.1, 35-47.
- Pfeifer, P. Deutsch S. (1980b). "Identification and interpretation of first order space-time ARMA models." *Technometrics*, Vol. 22, No.3, 397-408.
- Pfeifer, P. Deutsch S. (1981a). "Variance of the sample space-time correlation functions of contemporaneously correlated variables." *SIAM Journal of Applied Mathematics, Series A*, Vol. 40, No.1, 133-136.
- Pfeifer, P. Deutsch S. (1981b). "Seasonal space-time ARIMA modeling." *Geographical analysis*, Vol. 13, No. 2, 117-133.
- Pfeifer, P. Deutsch S. (1981c). "Space-time ARMA modeling with contemporaneously correlated innovations." *Technometrics*, Vol. 23, No. 4, 401-409.
- Pfeifer, P. Bodily, S. (1990). "A test of space-time ARMA modelling and forecasting of hotel data." *Journal of forecasting*, Vol. 8, No. 3, 255-272.
- Pumo, B. (1999). "Prediction of continuous time processes by C[0,1]-valued autoregressive process." *Stat. Inference Stoch. Process.*, Vol. 3, 1-13.
- Ruiz-Medina M. D. and Salmero R. (2010). "Functional Maximum likelihood estimation of ARH(P) models." *Stoch Environ Res Risk Assess*, Vol. 24, 131-146.
- Ruiz-Medina M. D. (2012). "spatial functional prediction from spatial autoregressive Hilbertian processes." *Environmetrics*, Vol. 23, 119-128.
- Soltani, A.R. and Hashemi, M. (2011). "Periodically correlated autoregressive Hilbertian." *Statistical Inference for Stochastic processes*, Vol. 14, No. 2, 177-188.
- Stoffer, D. (1986). "Estimation and interpretation of space-time ARMAX models in the presence of missing data." *Journal of the american statistical association*, Vol. 81, 762-772.



The inactivity time of the $(n - k + 1)$ -out-of- n system with exchangeable components under stress-strength models

Sh. Hashemi-Bosra^{1,2}, E. Salehi¹

¹Birjand University of Technology, Birjand, Iran

Abstract: In this paper, we consider a $(n - k + 1)$ -out-of- n system consisting of exchangeable components under stress-strength models. In following, the survival function of inactivity time of strengths of such systems is obtained. Finally, we present several properties of the mean inactivity time of a $(n - k + 1)$ -out-of- n system.

Keywords Stress-strength, Stochastic order, Mean inactivity time, Reliability.

Mathematics Subject Classification (2010): 62N05 60E15.

1 Introduction

A study of reliability of coherent system is one of the important subject that be investigated from various aspect. $(n - k + 1)$ -out-of- n systems are significant subclass of coherent systems that include series and parallel systems with $k = 1$ and $k = n$, respectively. In past years researchers have considered properties of mean inactivity time of coherent systems special $(n - k + 1)$ -out-of- n systems. The most important of that can be refer to [Asadi \(2006\)](#), [Khaledi and Shaked \(2007\)](#), [Navarro et al. \(2007\)](#), [Li and Zhang \(2008\)](#), [Li and Zhao \(2008\)](#), [Sadegh \(2008\)](#), [Tavangar and Asadi \(2010\)](#), [Goliforushani and Asadi \(2011\)](#), [Salehi and Asadi \(2012\)](#) and [Gupta \(2013\)](#). In the most studies researchers consider under the assumption that components of system are independent. However, in many practical situations, the components in a system are dependent because they share the same environment. In some situations, we encounter systems which have dependent exchangeable components. Recently,

²Speaker's Name: sh.hashemi1986@hotmail.com

many researchers have studied properties of mean inactivity time (MIT) function of system including dependent components. Among others, we refer to [Zhang \(2010\)](#), [Navarro and Rubio \(2011\)](#), [Sadegh \(2011\)](#), [Navarro and Rubio \(2012\)](#), [Rezapour et al. \(2013\)](#), [Tavangar and Asadi \(2014\)](#), [Salehi and Hashemi \(2014\)](#) and [Gupta et al. \(2015\)](#).

The stress-strength models that are very interested and important in the reliability theory, have been widely studied by several authors in the literatures. In these models, an unit is subjected to a random environmental stress over time. Let T and Y represent the random strength of the unit and the random stress placed on the unit, respectively. The reliability of unit is defined as the probability that the strength is larger than the stress. Thus the reliability of a unit is given by $R = P(T > Y)$. There have been numerators papers on inference about R under various assumptions. For a review of the topic we refer to [Kotz et al. \(2003\)](#), that is comprehensive reference on this area. Stress-strength models have also been discussed for a system consisting of several components. [Bhattacharyya and Johnson \(1974\)](#) paid their attention to apply stress-strength models for $(n - k + 1)$ -out-of- n system consisting of n components when the components placed under a common random stress Y and they could be obtained the reliability of the system under this setup as $P(T_{k:n} > Y)$, where T_i 's $i = 1, 2, \dots, n$ are the strengths of the components. [Hanagal \(2003\)](#) and [Eryilmaz \(2010\)](#) applied these models for the series and consecutive k -out-of- $n : G$ systems, respectively, and obtained some results. For more studies on this setup, we can refer to [Dewanji and Rao \(2001\)](#), [Eryilmaz \(2010\)](#), [Eryilmaz and Iscioglu \(2011\)](#) and [Bairamov et al. \(2015\)](#).

In the paper, we study the MIT function of components of a $(n - k + 1)$ -out-of- n system in the stress-strength setup when the lifetimes of components are exchangeable. We assume that the random stress and strengths are independent random variables and the random stress is common to all the components in the system level. We provide some results $(n - k + 1)$ -out-of- n under this setup, where T_1, T_2, \dots, T_n are the strengths of the components designed under the common stress. This paper is organized as follows. Some required concepts and preliminaries for presenting main results are presented in Section 2. In Section 3, we obtain the survival function of inactivity time of components of the $(n - k + 1)$ -out-of- n system consisting of exchangeable components under stress-strength model, and we give some stochastic ordering properties for this conditional variable.

2 Preliminaries

In this section we introduce some concepts and tools to deduce the main results in the next section.

Let T and Y be two independent random variable with distribution function F and G , respectively. Assume that we have a component with lifetime T that has failed at or some time before t , $t > 0$ under the applied stress Y . Consider the conditional random variable $\{t + Y - T|T \leq t + Y\}$ that is called inactivity time of component under stress Y . The MIT of the component under stress-strength models, can be defined as, for all t such that $F(t + y) > 0$,

$$\begin{aligned} \Psi^Y(t) &= E(t + Y - T|T \leq t + Y) \\ &= \frac{1}{\int_0^\infty F(t + y)dG_Y(y)} \int_0^\infty \int_0^\infty F(t + y - x)dG_Y(y)dx. \end{aligned}$$

Here, we give the concept of usual stochastic order, and for more details we refer the reader to Shaked and Shanthikumar (2007).

Definition 2.1. Let T and Z be two nonnegative random variables with survival functions \bar{F} and \bar{G} , respectively. T is said to be smaller than Z in the usual stochastic order, denoted by $T \leq_{st} Z$, if for all t , $\bar{F}(t) \leq \bar{G}(t)$.

In the following, we present the reliability function of the inactivity time of components of $(n - k + 1)$ -out-of- n systems, that is obtained by Tavangar and Asadi (2014).

Lemma 2.2. Let T_1, T_2, \dots, T_n be the exchangeable lifetimes of components. The reliability function of $\{t - T_{r:n}|T_{k:n} \leq t\}$, denoted by $\psi_{r:n}^k(x|t)$, is equal to for $1 \leq r < k \leq n$ and $x, t > 0$,

$$\psi_{r:n}^k(x|t) = \frac{\sum_{i=k}^n \binom{n}{i} \sum_{j=r}^i \binom{i}{j} p_{i,j,n}(t, x)}{\sum_{i=k}^n \binom{n}{i} p_{i,n}(t)},$$

where

$$p_{i,j,n}(x, t) = \sum_{l=0}^j \sum_{v=0}^{i-j} \binom{j}{l} \binom{i-j}{v} (-1)^{l+v} \bar{F}(\underbrace{t-x, \dots, t-x}_{i-j+l-v}, \underbrace{t, \dots, t}_{n-i+v}, \underbrace{0, \dots, 0}_{j-l}), \tag{2.1}$$

and

$$p_{i,n}(t) = \sum_{j=0}^i \binom{i}{j} (-1)^j \bar{F}(\underbrace{t, \dots, t}_{n-i+j}, \underbrace{0, \dots, 0}_{i-j}). \tag{2.2}$$

3 Main results

Let T_1, T_2, \dots, T_n be the exchangeable strengths of the components of the $(n - k + 1)$ -out-of- n system, and let the random variable Y be the common stress which is independent of T_1, T_2, \dots, T_n . Under condition that the k th strongest strength component fails at the strength level t , the inactivity strength of r th strongest components as defines as

$$T_{k:n}^{r,Y}(t) = \{t + Y - T_{r:n} | T_{k:n} \leq t + Y\}.$$

The mean inactivity time (MIT) of the failed components under the stress-strength model, denoted by $\Psi_{k:n}^{r,Y}(t)$, can be defined as

$$\Psi_{k:n}^{r,Y}(t) = E(t + Y - T_{r:n} | T_{k:n} \leq t + Y), \quad \text{for all } t > 0. \quad (3.1)$$

Now, we have the following theorem.

Theorem 3.1. *Let T_1, T_2, \dots, T_n be the exchangeable strengths of components. The reliability function of the inactivity time of components for the $(n - k + 1)$ -out-of- n system under the stress Y , denoted by $\psi_{k:n}^{r,Y}(x|t)$, is equal to for $1 \leq r < k \leq n$ and $x, t > 0$,*

$$\psi_{k:n}^{r,Y}(x|t) = \frac{\int_0^\infty \sum_{i=k}^n \binom{n}{i} \sum_{j=r}^i \binom{i}{j} p_{i,j,n}(x, t + y) dG_Y(y)}{\int_0^\infty \sum_{i=k}^n \binom{n}{i} p_{i,n}(t + y) dG_Y(y)}, \quad (3.2)$$

where $p_{i,j}(x, t + y)$ and $p_{i,n}(t + y)$ is defined as (2.1) and (2.2), respectively.

Proof. The proof follows from Lemma 2.2. □

If the MIT function of components of the $(n - k + 1)$ -out-of- n system under the stress Y , denoted by $\Psi_{k:n}^{r,Y}(t)$, is equal to

$$\Psi_{k:n}^{r,Y}(t) = \int_0^\infty \psi_{k:n}^{r,Y}(x|t) dx.$$

Corollary 3.2. *Let T_1, T_2, \dots, T_n be exchangeable random variables under stress Y . For all $x, t > 0$, we have*

$$P(t + Y - T_1 | T_{n:n} \leq Y + t) = \frac{1}{n} \sum_{r=1}^n \int_0^\infty P(t + Y - T_{r:n} | T_{n:n} \leq Y + t) dG_Y(y).$$

Proof. By using the corresponding properties for order statistics the proof is immediate. \square

The next proposition shows that under the stress-strength model for any fixed k , n and t , the inactivity time of components of the $(n - k + 1)$ -out-of- n system is stochastically decreasing in r .

Proposition 3.3. *Let T_1, T_2, \dots, T_n be exchangeable strengths and Y be a random stress. For $t > 0$, and $1 \leq r < k < n$, we have*

$$T_{k,n}^{r+1,Y}(t) \leq_{st} T_{k,n}^{r,Y}(t).$$

Proof. By using (3.2) in Theorem 3.1, the proof can be obtained easily. \square

Theorem 3.4. *Let $\mathbf{T} = (T_1, T_2, \dots, T_n)$ be exchangeable strengths and Y is a random stress. Then for $t > 0$, and $1 \leq l \leq r \leq k \leq n - 1$, we have*

$$\Psi_{k:n-1}^{l,Y}(t) \geq \frac{n-k}{n} \Psi_{k:n}^{r,Y}(t).$$

Proof. From Eq. (3.4.5) in David and Nagaraja (2003) (2003, p. 48), i.e

$$\begin{aligned} nP(T_{r:n-1} \leq Y + t - x, T_{k:n-1} \leq t + Y) &= rP(T_{r+1:n} \leq Y + t - x, T_{k+1:n} \leq Y + t) \\ &+ (k-r)P(T_{r:n} \leq Y + t - x, T_{k+1:n} \leq Y + t) + (n-k)P(T_{r:n} \leq Y + t - x, T_{k:n} \leq Y + t), \end{aligned}$$

and from $F_{k:n}(y+t) \geq F_{k:n-1}(y+t)$ for $t > 0$, it is easy to show that for $x, t > 0$,

$$P(t+Y - T_{r:n-1} > x | T_{k:n-1} \leq Y+t) \geq \frac{n-k}{n} P(t+Y - T_{r:n} > x | T_{k:n} \leq Y+t).$$

Also, from Propositions 3.3, we can write for $x, t > 0$ and $1 \leq l \leq r \leq k < n$,

$$P(t+Y - T_{l:n-1} > x | T_{k:n-1} \leq Y+t) \geq \frac{n-k}{n} P(t+Y - T_{r:n} > x | T_{k:n} \leq Y+t).$$

Finally, twice integrating the previous expression from 0 to ∞ respect to x and then y , the result holds. \square

Example 3.5. *Let the exchangeable lifetimes of components $\mathbf{T} = (T_1, T_2, \dots, T_n)$ have a FGM multivariate exponential distribution with parameter λ_1 under the exponential stress Y with parameter λ_2 .*

Also assume that Y is independent of \mathbf{T} . In special case, let $\theta_{12} = \theta_{13} = \dots = \theta_{12\dots n} = \theta$. In this case for $n = 3$, $\theta \in [-1/3, 1/2]$, and then using (3.2) in Theorem 3.1, we have

$$\begin{aligned}\psi_{2:2}^{1,Y}(x|t) &= \frac{\int_0^\infty \left(2F(t+y-x, t+y) - F(t+y-x, t+y-x)\right) dG_Y(y)}{\int_0^\infty F(t+y, t+y) dG_Y(y)} \\ \psi_{2:3}^{2,Y}(x|t) &= \frac{\int_0^\infty \left(3F(t+y-x, t+y-x) - 2F(t+y-x, t+y-x, t+y-x)\right) dG_Y(y)}{\int_0^\infty \left(3F(t+y, t+y) - 2F(t+y, t+y, t+y)\right) dG_Y(y)}.\end{aligned}$$

Hence, from Theorem 3.4, it can be shown that $\Psi_{2:2}^{1,Y}(t) \geq \frac{1}{3}\Psi_{2:3}^{2,Y}(t)$. The graphs of $\Psi_{2:2}^{1,Y}(t)$ and $\frac{1}{3}\Psi_{2:3}^{2,Y}(t)$, for $\theta = 0.5$, $\lambda_1 = 0.5$ and $\lambda_2 = 0.75$, are plotted in Figure 3.

Proposition 3.6. Let (T_1, T_2, \dots, T_n) are exchangeable strengths and Y is a random stress and T_{n+1} is independent of the (T_1, T_2, \dots, T_n) . Then for $x, t > 0$ we have

$$E(t+Y - T_{r:n} > x | T_{n+1:n+1} \leq Y+t) \geq \frac{1}{n+1} E(t+Y - T_{r:n+1} | T_{n:n+1} \leq Y+t).$$

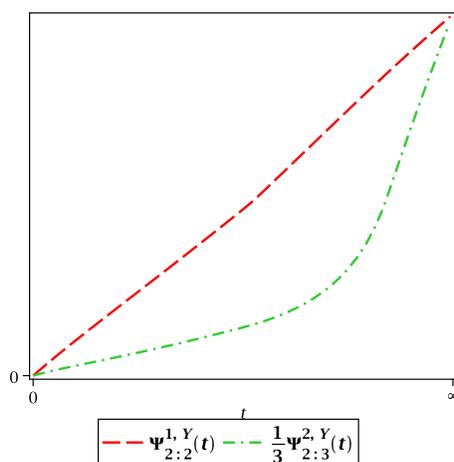
Proof. It can be easily shown that $\{t+Y - T_{r:n} | T_{n+1:n+1} \leq t+Y\} \stackrel{st}{=} \{t+Y - T_{r:n} | T_{n:n} \leq Y+t\}$, and hence

$$E(t+Y - T_{r:n} > x | T_{n+1:n+1} \leq Y+t) = E(t+Y - T_{r:n} | T_{n:n} \leq Y+t).$$

Hence, by replacing $k = n$ in Theorem 3.4, the proof is complete. \square

References

- Asadi, M. (2006). On the mean past lifetime of the components of a parallel system. *J. Statist. Plann. Infer.*, **136**, 1197-1206.
- Bairamov, I.G., Gurler, S. and Ucer, B.(2015). On the Mean Remaining Strength of the k -Out-of- n :F System with Exchangeable Components. *Commun. Statist. Simul. Comput.*, **1**, 119-132.
- Bhattacharyya, G. K. and Johnson, R. A. (1974). Estimation of reliability in a multicomponent stress-strength model. *J. Amer. Statist. Assoc.* **69**, 966-970.



David, H. A., and Nagaraja, H. N. A., (2003). *Order statistics*. John Wiley & Sons, NewYork.

Dewanji, A., and Rao, T.S. (2001). On system reliability under stress-strength modeling. *Commun. Statist. Theor. Meth.*, **30**, 1185-1196.

Eryilmaz, S. (2010). On system reliability in stress-strength setup. *Statis. Probab. Lett.*, **80**, 834-839.

Eryilmaz, S., and Iscioglu, F. (2011). Reliability evaluation for a multi-state system under stress-strength setup. *Commun. Statist. Theor. Meth.*, **40**, 547-558.

Goliforushani, S. and Asadi, M. (2011). Stochastic ordering among inactivity times of coherent systems. *Sankhya*, **73**, 241-262.

Gupta, N. (2013). Stochastic comparison of residual life times and inactivity times of coherent systems. *J. Appl. Probab.*, **30**, 848-860.

Gupta, N., Misra, N., and Kumar, S. (2015). Stochastic comparisons of residual lifetimes and inactivity times of coherent systems with dependent identically distributed components. *Europ. J. Oper. Res.*, **240**, 425-430.

- Hanagal, D. (2003). Estimation of system reliability in multicomponent series stress-strength models. *J. Ind. Statist. Assoc.* **41**, 1-7.
- Khaledi, B-E. and Shaked, M. (2007). Ordering conditional lifetimes of coherent systems. *J. Statist. Plann. Infer.*, **137**, 1173-1184.
- Kotz, S., Lumelskii, Y. and Pensky, M. (2003). *The Stress-Strength Model and its Generalizations*. Singapore: World Scientific.
- Li, X. and Zhang, Z. (2008). Some stochastic comparisons of conditional coherent systems. *J. Appl. Stoch. Model. Bus. Ind.*, **24**, 541-549.
- Li, X. and Zhao, P., (2008). Stochastic comparison on general inactivity time and general residual life of k -out-of- n systems. *Commun. Statist. Simul. Comput.*, **37**, 1005-1019.
- Navarro, J. and Rubio, R. (2011). A note on necessary and sufficient conditions for ordering properties of coherent systems with exchangeable components. *Naval Res. Logist.* **58**, 478-489.
- Navarro, J. and Rubio, R. (2012). Comparisons of coherent systems with nonidentically distributed components. *J. Statist. Plann. Infer.*, **142**, 1310-1319.
- Navarro, J., Ruiz, J. M. and Sandoval, C. J., (2007). Properties of coherent systems with dependent components. *Commun. Statist. Theor. Meth.*, **36**, 1-17.
- Rezapour, M., Salehi, E. T. and Alamatsaz, M. H., (2013), Stochastic comparison of residual and past lifetimes of $(n - k + 1)$ -out-of- n systems with dependent components, *Commun. Statist. Theor. Meth.*, **42**, 2158-2199.
- Sadegh, M. K., (2008). Mean past and mean residual life of parallel system with nonidentical components. *Commun. Statist. Theor. Meth.*, **37**, 1137-1145.
- Sadegh, M. K., (2011). A note on the mean residual life function of a coherent system with exchangeable or nonidentical components. *J. Statist. Plann. Infer.*, **141**, 3267-3275.

- Salehi, E. T. and Asadi, M., (2012). Results on the past lifetime of $(n - k + 1)$ -out-of- n structures with nonidentical components. *Metrika*, **75**, 439-454.
- Salehi, E. and Hashemi, M., (2014). The mean general residual lifetime of $(n - k + 1)$ -out-of- n systems including exchangeable components. *Submitted*.
- Tavangar, M. and Asadi, M., (2010). A study on the mean past lifetime of the components of $(n - k + 1)$ -out-of- n system at the system level. *Metrika*, **72**, 59-73.
- Tavangar, M. and Asadi, M. (2014). The Inactivity Time of Exchangeable Components of k -out-of- n Structures. *Sankhya*, **136**, 1197-1206.
- Zhang, Z., (2010). Ordering conditional general coherent systems with exchangeable components. *J. Statist. Plann. Infer.*, **140**, 454-460.
- Zhao, P., Li, X. and Balakrishna, N., (2008). Conditional ordering of k -out-of- n systems with independent but non-identical components. *J. Appl. Probab.*, **45**, 1113-1125.



Adaptive Progressively Type-II Censored Competing Risks Data

Fariba Hemmati¹, Esmail Khorram²

¹Central Bank of The Islamic Republic of Iran

²Amirkabir University of Technology, Tehran, Iran.

Abstract: In this paper, a competing risks model based on exponential distributions is considered under the adaptive Type-II progressively censoring scheme, for life testing or reliability experiment. Moreover, we assumed that some causes of failures are unknown. The maximum likelihood estimators (MLEs) of unknown parameters are established. The asymptotic distributions of the obtained estimators are derived to construct the confidence intervals as well as the two different bootstraps of different unknown parameters. Under suitable priors on the unknown parameters, Bayes estimates and the corresponding two sides of Bayesian probability intervals are obtained. Also, for the purpose of evaluating the average bias and mean square error of the MLEs, and comparing the confidence intervals based on all mentioned methods, a simulation study was carried out.

Keywords Competing risks; Hybrid censoring; Type-II progressive censoring.

Mathematics Subject Classification (2010): 62N01.

1 Introduction

Childs et al. (2008) considered the progressive Type-II hybrid censoring (PHC) scheme in which n units are placed on a test with censoring scheme (R_1, R_2, \dots, R_m) and stopping time $T^* = \min\{X_{m:m:n}, T\}$, where $X_{1:m:n} \leq X_{2:m:n} \leq \dots \leq X_{m:m:n}$ are the order observed failure times resulting from the progressively censored experiment and T is fixed in advance. Although, in order to control the total time on test, the experiment time is fixed by the experimenter, so less than m failures (or even equal to

¹Fariba Hemmati : hemati@aut.ac.ir

zero) may be observed which have an adverse effect on the efficiency of the inferential producer based on the progressive Type-II hybrid censoring scheme. For the purpose of increasing the efficiency of statistical analysis as well as saving the total test time, [Ng et al. \(2009\)](#) introduced an adjustment of progressive Type-II hybrid censoring scheme, so called adaptive progressive Type-II censoring (APC) scheme. Under this scheme, the number of observed failures m is fixed in advance but the experimental time is allowed to run over the (pre-fixed) threshold time $T > 0$. If $X_{m:m:n} < T$, the experiment stops at time $X_{m:m:n}$, and we will have a usual progressive Type-II censoring scheme with the pre-fixed progressive censoring scheme (R_1, R_2, \dots, R_m) . If $X_{J:m:n} < T < X_{J+1:m:n}$, where $J + 1 < m$, we adapt the number of items progressively removed from the experiment upon failure by setting $R_{J+1} = R_{J+2} = \dots = R_{m-1} = 0$ and $R_m = n - m - \sum_{i=1}^J R_i$. Thus, the effectively applied scheme is $(R_1, \dots, R_J, 0, \dots, 0, n - m - \sum_{i=1}^J R_i)$, where $J = \max\{j : X_{j:m:n} < T\}$, that is, the first observed failure time exceeding the ideal total time T . Put another way, as long as the failures occur before time T , the initially planned progressive scheme is applied. After passing time T , we do not withdraw any items at all except for the time of the m -th failure where all remaining surviving items are removed. This determination results in terminating the experiment as soon as the $(J + 1)$ -th failure time is greater than T , and the total test time will not be too far away from time T . If $T = 0$, the scheme will lead us to the case of the conventional Type-II censoring scheme, and if $T \rightarrow \infty$, we will have a usual progressive Type-II censoring scheme. This approach illustrates how an experimenter can control the experiment. The experimenter can decide to change the value of T as a compromise between a shorter experimental time and a higher chance to observe extreme failures (You can see [Lin et al. \(2009\)](#) and [Hemmati and Khorram \(2013\)](#)). Furthermore, in reliability analysis, a failure is commonly associated with one of several fatal risk factors to which the test unit is exposed. Since it is not usually possible to study the test units with an isolated risk factor, it requires assessing each risk factor in the presence of other risk factors. In order to analyze such a competing risks model, each failure observation needs to be expressed by a bivariate form composed of a failure time and the cause of failure (See [Kundu and Joarder \(2006b\)](#)). In certain situations, it is observed that the determination of the cause of failure may be very expensive or difficult to obtain. In those conditions, one might observe the failure time, but not the corresponding cause of failure. In this paper, we discuss the adaptive progressively Type-II

censoring (APC) scheme when the life time of the different risk factors are independently exponentially distributed and the number of failure causes is unknown.

2 Model description and Maximum Likelihood Estimators

Consider a lifetime experiment with n identical units where its lifetimes are described by identically independent and distributed (i.i.d) random variables X_1, \dots, X_n . Without loss of generality, we assume that there are only two causes of failure. We have $X_i = \min\{X_{1i}, X_{2i}\}$ for $i = 1, \dots, n$, where X_{hi} , $h = 1, 2$, denotes the latent failure time of the i th unit under the h th cause of failure. Also, stochastically independent for $i = 1, \dots, n$, the latent failure times X_{hi} , $h = 1, 2$, follow the exponential distribution and the pairs (X_{1i}, X_{2i}) , $i = 1, \dots, n$ are i.i.d., such that the cumulative distribution function of X_{hi} is: $F_h(x) = 1 - e^{-\lambda_h x}$, $x > 0$. In the presence of competing risks, the data from an adaptive Type-II progressively censoring scheme is as follows: $(X_{1:m:n}, \delta_1, R_1), \dots, (X_{J:m:n}, \delta_J, R_J), (X_{J+1:m:n}, \delta_{J+1:m:n}, 0), \dots, (X_{m-1:m:n}, \delta_{m-1:m:n}, 0), (X_{m:m:n}, \delta_m, R_m)$, where $J = \max\{j : X_{j:m:n} < T\}$, $R_m = n - m - \sum_{i=1}^m R_i$ and $\delta_i \in \{1, 2, *\}$. Here, $\delta_i = h$, $h = 1, 2$ means that unit i has failed at time $X_{i:m:n}$ due to cause h , while $\delta_i = *$ indicates unidentifying the cause of unit i to fail. Let $I(A)$ is the indicator of the event A , then, the random variables $m_1 = \sum_{i=1}^m I(\delta_i = 1)$ and $m_2 = \sum_{i=1}^m I(\delta_i = 2)$ describe the number of failures due to the first and second cause of failures, respectively. Also, $m_3 = \sum_{i=1}^m I(\delta_i = *)$ is the number of failures having failure times but corresponding causes of failure are unknown. We denote $m_1 + m_2 = m^*$ and therefore, $m^* + m_3 = m$, in which m^* is considered positive and fixed. Using the independence of the latent failure times X_{1i} and X_{2i} , $i = 1, \dots, n$, we obtain the probabilities $\pi_h = P(X_{hi} \leq X_{(3-h)i}) = \frac{\lambda_h}{\lambda_h + \lambda_{3-h}}$, $h = 1, 2$, as the relative risk due to the h th cause. Notice that both m_1 and m_2 follow binomial distributions with sample size m^* . Hence, $m_h \sim \text{bin}(m^*, \frac{\lambda_h}{\lambda_h + \lambda_{3-h}})$, $h = 1, 2$. Given censoring scheme $\mathcal{R} = (R_1, R_2, \dots, R_J, 0, \dots, 0, n - m - \sum_{i=1}^J R_i)$, where $J = \max\{j : X_{j:m:n} < T\}$, the likelihood function of the observed data $(x_1, \delta_1), \dots, (x_m, \delta_m)$ is expressed by $L(\text{data}; \lambda_1, \lambda_2 | J = j) = c(\mathcal{R}) \prod_{i=1}^m \left\{ \left[f_1(x_i) \bar{F}_2(x_i) \right]^{I(\delta_i=1)} \left[f_2(x_i) \bar{F}_1(x_i) \right]^{I(\delta_i=2)} \right.$
 $\times \left. \left[f_1(x_i) \bar{F}_2(x_i) + f_2(x_i) \bar{F}_1(x_i) \right]^{I(\delta_i=*)} \right\} \times \prod_{i=1}^J \left\{ \left[\bar{F}_1(x_i) \bar{F}_2(x_i) \right]^{R_i} \right\} \times \left[\bar{F}_1(x_m) \bar{F}_2(x_m) \right]^{(n-m-\sum_{i=1}^J R_i)}$, where $\bar{F}_h(\cdot) = 1 - F_h(\cdot)$, $h = 1, 2$, and $c(\mathcal{R}) = n(n-1-R_1)(n-2-R_1-R_2) \dots (n-j+1-R_1-\dots-$

$R_j) \times (n - j + 1 - R_1 - \dots - R_j) \cdots (n - m + 1 - R_1 - \dots - R_j)$, is the normalizing constant. We get $L(\text{data}; \lambda_1, \lambda_2 | J = j) = c(\mathcal{R}) \lambda_1^{m_1} \lambda_2^{m_2} (\lambda_1 + \lambda_2)^{(m - m^*)} \exp \left\{ -(\lambda_1 + \lambda_2) \left[\sum_{i=1}^m x_i + \sum_{i=1}^j R_i x_i + (n - m - \sum_{i=1}^j R_i) x_m \right] \right\}$. Taking the logarithm of $L(\text{data}; \lambda_1, \lambda_2 | J = j)$ and ignoring the additive constant, we obtain $\mathcal{L} = m_1 \log \lambda_1 + m_2 \log \lambda_2 + (m - m^*) \log(\lambda_1 + \lambda_2) - (\lambda_1 + \lambda_2) \gamma(\mathbf{x}, j)$ where

$$\gamma(\mathbf{x}, j) = \sum_{i=1}^m x_{i:m:n} + \sum_{i=1}^j R_i x_{i:m:n} + (n - m - \sum_{i=1}^j R_i) x_{m:m:n}. \quad (2.1)$$

Equating the first derivations of \mathcal{L} to zero, we get the MLEs as in the following form

$$\hat{\lambda}_h = \frac{m m_h}{m^* \gamma(\mathbf{x}, j)}, h = 1, 2. \quad (2.2)$$

3 Bayesian estimation

In this section, we provide the Bayes estimators and the corresponding credible intervals of the unknown parameters λ_1 and λ_2 under the square error loss function. To do that, the following additional assumptions are required:

(i) The parameters λ_1 and λ_2 are considered as independent random variables.

(ii) The random variable λ_h , $h = 1, 2$, has gamma prior distribution with known shape and scale parameters α_h and β_h in the following form $\pi_h(\lambda_h) = \frac{\beta_h^{\alpha_h}}{\Gamma(\alpha_h)} \lambda_h^{\alpha_h - 1} e^{-\beta_h \lambda_h}$, $\lambda_h > 0$. Using binomial expansion, the likelihood function can be written as

$$L(\text{data}; \lambda_1, \lambda_2 | J = j) = c(\mathcal{R}) \exp \left\{ -(\lambda_1 + \lambda_2) \gamma(\mathbf{x}, j) \right\} \sum_{i=0}^{m - m^*} \binom{m - m^*}{i} \lambda_1^{m_1 + i} \lambda_2^{m_2 + m - m^* - i}. \quad (3.1)$$

Theorem 3.1. Under the assumptions (i) and (ii), the marginal posterior PDF's of λ_1 and λ_2 are given, respectively, by

$$\pi_1(\lambda_1 | \text{data}, J = j) = \frac{1}{A} \exp \{ -(\beta_1 + \gamma(\mathbf{x}, j)) \lambda_1 \} \times \sum_{i=0}^{m - m^*} \binom{m - m^*}{i} \frac{\Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_2 + \gamma(\mathbf{x}, j))^{\alpha_2 + m_2 + m - m^* - i}} \lambda_1^{\alpha_1 + m_1 + i - 1},$$

and

$$\pi_2(\lambda_2 | \text{data}, J = j) = \exp \{ -(\beta_2 + \gamma(\mathbf{x}, j)) \lambda_2 \} \times \frac{1}{A} \sum_{i=0}^{m - m^*} \binom{m - m^*}{i} \frac{\Gamma(\alpha_1 + m_1 + i)}{(\beta_1 + \gamma(\mathbf{x}, j))^{\alpha_1 + m_1 + i}} \lambda_2^{\alpha_2 + m_2 + m - m^* - i},$$

where $0 < \lambda_1, \lambda_2 < \infty$, and $A = \sum_{i=0}^{m - m^*} \binom{m - m^*}{i} \frac{\Gamma(\alpha_1 + m_1 + i) \Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(\mathbf{x}, j))^{\alpha_1 + m_1 + i} (\beta_2 + \gamma(\mathbf{x}, j))^{\alpha_2 + m_2 + m - m^* - i}}$.

Theorem 3.2. Under the assumptions (i) and (ii) the Bayes estimators for λ_1 and λ_2 are given, respectively, by

$$\hat{\lambda}_1 = \frac{1}{A} \sum_{i=0}^{m-m^*} \binom{m-m^*}{i} \frac{\Gamma(\alpha_1 + m_1 + i + 1) \Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(\mathbf{x}, j))^{\alpha_1 + m_1 + i + 1} (\beta_2 + \gamma(\mathbf{x}, j))^{\alpha_2 + m_2 + m - m^* - i}}, \text{ and} \quad (3.2)$$

$$\hat{\lambda}_2 = \frac{1}{A} \sum_{i=0}^{m-m^*} \binom{m-m^*}{i} \frac{\Gamma(\alpha_1 + m_1 + i) \Gamma(\alpha_2 + m_2 + m - m^* - i + 1)}{(\beta_1 + \gamma(\mathbf{x}, j))^{\alpha_1 + m_1 + i} (\beta_2 + \gamma(\mathbf{x}, j))^{\alpha_2 + m_2 + m - m^* - i + 1}}, \quad (3.3)$$

4 Confidence Intervals

Approximate Confidence Intervals We present an approximate method to construct CIs for λ_h , $h = 1, 2$, using the properties of the MLEs for large sample sizes. Under the usual regularity conditions, the MLEs of the parameters λ_h , $h = 1, 2$, are asymptotically normal and efficient. As $m_1 \rightarrow \infty$, $m_2 \rightarrow \infty$, and $m \rightarrow \infty$, then the expected Fisher information matrix of the parameters λ_1 and λ_2 is

$$I(\lambda_1, \lambda_2) = \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix} = \frac{1}{(\lambda_1 + \lambda_2)^2} \begin{bmatrix} \frac{m^* \lambda_2 + m \lambda_1}{\lambda_1} & m - m^* \\ m - m^* & \frac{m^* \lambda_1 + m \lambda_2}{\lambda_2} \end{bmatrix}, \quad (4.1)$$

where the elements of the expected Fisher information matrix of the parameters λ_1 and λ_2 , $I_{ik}(\lambda_1, \lambda_2)$, $i, k = 1, 2$, can be obtained by $I_{ik}(\lambda_1, \lambda_2) = -E\left\{\frac{\partial^2 \mathcal{L}}{\partial \lambda_i \partial \lambda_k}\right\}$. Substituting $\hat{\lambda}_h$ for λ_h , we obtain the observed Fisher information matrix and the asymptotic variances of $V_{hh} = \text{var}(\hat{\lambda}_h) = I_{hh}^{-1}(\hat{\lambda}_1, \hat{\lambda}_2)$, $h = 1, 2$. Therefore, by using the asymptotic normality of the MLEs, the two sided $100(1-\alpha)\%$ approximate confidence intervals for $\hat{\lambda}_h$, $h = 1, 2$ is given by $(\max\{0, \hat{\lambda}_h - z_{\frac{\alpha}{2}} \sqrt{V_{hh}}\}, \hat{\lambda}_h + z_{\frac{\alpha}{2}} \sqrt{V_{hh}})$, $h = 1, 2$, where $Z_{\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ th quantile of a standard normal distribution.

Two sided Bayesian credible intervals The $100(1-\alpha)\%$ two sided Bayesian probability intervals of λ_h , $h = 1, 2$, say (u_h, v_h) , can be derived by solving the following two equations, with respect to u_h and v_h

$$\frac{\alpha}{2} = \int_0^{u_h} \pi_h(\lambda_h | \text{data}, J = j) d\lambda_h, \quad (4.2)$$

$$1 - \frac{\alpha}{2} = \int_0^{v_h} \pi_h(\lambda_h | \text{data}, J = j) d\lambda_h. \quad (4.3)$$

Substituting (3.2) into (??) and (4.3), the $100(1 - \alpha)\%$ two sided Bayesian probability interval of λ_1 can be obtained as the solutions of the following equations with respect to u_1 and v_1 :

$$\frac{\alpha}{2} = \frac{1}{A} \sum_{i=0}^{m-m^*} A_i \Gamma(\alpha_1 + m_1 + i, (\beta_1 + \gamma(\mathbf{x}, j))u_1), \quad (4.4)$$

$$1 - \frac{\alpha}{2} = \frac{1}{A} \sum_{i=0}^{m-m^*} A_i \Gamma(\alpha_1 + m_1 + i, (\beta_1 + \gamma(\mathbf{x}, j))v_1), \quad (4.5)$$

$A_i = \frac{\binom{m-m^*}{i} \Gamma(\alpha_1 + m_1 + i) \Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(\mathbf{x}, j))^{\alpha_1 + m_1 + i} (\beta_2 + \gamma(\mathbf{x}, j))^{\alpha_2 + m_2 + m - m^* - i}}$, and $\Gamma(a, x)$ is the incomplete gamma function defined by $\Gamma(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt$. Equations (4.4) and (4.5) do not yield explicit solutions for u_1 and v_1 and have to be solved numerically to obtain (u_1, v_1) . If $m - m^* = 0$, the $100(1 - \alpha)\%$ two sided Bayesian probability interval of λ_h , $h = 1, 2$ can be obtained using gamma distribution, as follows

$$\frac{\alpha}{2} = \int_0^{u_h} \frac{(\beta_h + \gamma(\mathbf{x}, j))^{\alpha_h + m_h}}{\Gamma(\alpha_h + m_h)} \lambda_h^{\alpha_h + m_h - 1} e^{-(\beta_h + \gamma(\mathbf{x}, j))} d\lambda_h, \quad (4.6)$$

$$1 - \frac{\alpha}{2} = \int_0^{v_h} \frac{(\beta_h + \gamma(\mathbf{x}, j))^{\alpha_h + m_h}}{\Gamma(\alpha_h + m_h)} \lambda_h^{\alpha_h + m_h - 1} e^{-(\beta_h + \gamma(\mathbf{x}, j))} d\lambda_h, \quad (4.7)$$

Bootstrap Confidence Intervals In this subsection, we use two confidence intervals based on bootstrapping. (a) The percentile bootstrap (Boot-p) confidence interval of Efron (1982), (b) The bootstrap-t (Boot-t) confidence interval of Hall (1988). To find these Bootstrap intervals, in the first step, we generate original samples from adaptive Type-II progressively censored competing risk data. Next, we estimate $\hat{\lambda}_h$, $h = 1, 2$, from the sample using (2.2) as before. In the second step, we generate a bootstrap sample $\{(X_{1:m:n}^*, \delta_1^*), \dots, (X_{m:m:n}^*, \delta_m^*)\}$, using $\hat{\lambda}_1$ and $\hat{\lambda}_2$, R_1, \dots, R_m and T , obtain the bootstrap estimate of $\hat{\lambda}_h$, say $\hat{\lambda}_h^*$, $h = 1, 2$. In the third step, repeat the second step N-Boot times.

Boot-p: Let $\widehat{CDF}(x) = P(\hat{\lambda}_h^* \leq x)$, be the cumulative distribution function of $\hat{\lambda}_h^*$, $h = 1, 2$. For a given x define $\hat{\lambda}_{h\text{Boot-p}}(x) = \widehat{CDF}^{-1}(x)$, then the approximate $100(1 - \alpha)\%$ confidence interval for $\hat{\lambda}_h$, $h = 1, 2$, is indicated by $(\hat{\lambda}_{h\text{Boot-p}}(\frac{\alpha}{2}), \hat{\lambda}_{h\text{Boot-p}}(1 - \frac{\alpha}{2}))$.

Boot-t: After generating the bootstrap samples in the second step and calculating $\hat{\lambda}_h$, we need to use them to compute the estimate of $V(\hat{\lambda}_h^*)$ from the observed Fisher information matrix in section 4.1. Then, we determine the statistic T^* given by $T^* = \frac{\sqrt{m_h^*}(\hat{\lambda}_h^* - \hat{\lambda}_h)}{\sqrt{V(\hat{\lambda}_h^*)}}$, Now, suppose $\widehat{CDF}(x) = P(T^* \leq x)$, be the cumulative distribution function of T^* . For a given x , define $\hat{\lambda}_{h\text{Boot-t}}(x) =$

$\hat{\lambda}_h + m_h^{-\frac{1}{2}} \sqrt{\hat{V}(\hat{\lambda}_h^*) \widehat{CDF}^{-1}(x)}$, then the approximate $100(1 - \alpha)\%$ confidence interval for $\hat{\lambda}_h$, $h = 1, 2$, is presented by $(\hat{\lambda}_{hBoot-t}(\frac{\alpha}{2}), \hat{\lambda}_{hBoot-t}(1 - \frac{\alpha}{2}))$.

5 Simulation study

In this section, we report the obtained results of a simulation study carried out by software R, to illustrate the theoretical results obtained in the previous sections. This simulation has been done by considering different values of n , m and T , and by choosing $\lambda_1 = 0.6$ and $\lambda_2 = 0.8$ in all the cases. We have used three different sampling schemes as follows: Scheme 1 : $R_1 = \dots = R_{m-1} = 0$ and $R_m = n - m$, Scheme 2 : $R_1 = n - m$ and $R_2 = \dots = R_m = 0$, Scheme 3 : $R_1 = \dots = R_{m-1} = 1$ and $R_m = n - 2m + 1$. For each case, the MLEs and Bayes estimators as described before they are computed based on 3000 simulations. It is assumed that 10% of the data have the unknown cause of failure, and for Bayes estimation, parameters λ_1 and λ_2 are treated as random variables with gamma prior distributions with parameters (3.6,6.0) and (6,6.4), respectively. Based on Tables 1-2, it can be seen that the absolute biases and MSEs of the MLEs based on the APC scheme are always smaller than those based on the PHC scheme. For small Ts, the difference between the absolute biases and MSEs of two schemes is more sensible. Therefore, the APC scheme gives better performance in parameter estimation. Comparing Table 2 and 3, we observed that the average biases and MSEs of the Bayes estimators are always smaller than those based on the MLEs. Therefore, we would recommend using the Bayesian method in parameter estimation. The length of the different confidence intervals and the corresponding coverage probabilities are reported in Table 4. A comparison in terms of interval estimation, approximate, Boot-p, Boot-t confidence intervals and Bayes credible intervals are quite satisfied, and all of them are able to keep the nominal coverage probabilities, even for small sample sizes. The Bayes credible intervals provides a smaller width and larger coverage probabilities in comparison to other methods.

6 Conclusion

In this article, we have discussed a competing risks model from exponential distribution when the data are the adaptive Type-II progressive censoring and also the fixed number of causes of failure is unknown and, derived the MLEs of the parameters, Bayes estimates and credible intervals under the assumption of gamma priors, and confidence intervals for the parameters.

References

- Childs, A., Chandrasekar, B., Balakrishnan, N. (2008). Exact likelihood inference for an exponential parameter under progressive hybrid censoring schemes. *In: Vonta, F., Nikulin, M., Limnios, N., Huber-Carol, C., eds. Statistical Models and Methods for Biomedical and Technical Systems. Boston: Birkhauser:323-334.*
- Hemmati, F., Khorram, E. (2013). Statistical analysis of the log-normal distribution under Type-II progressive hybrid censoring schemes. *Communications in Statistics-Simulation and Computation.* 42:1, 52-75.
- Kundu, D., Joarder, A. (2006). *Analysis of Type II progressively hybrid censored competing risks data.* Journal of Modern Applied Statistical Methods. 5:186-204.
- Lin, C. T., Ng, H. K. T., Chan, P. S. (2009). Statistical inference of Type-II progressively hybrid censored data with Weibull lifetimes. *Communications in Statistics-Theory and Methods.* 38:1710-1729.
- Ng, H. K. T., Kundu, D., Chan, P. S. (2009). Statistical analysis of exponential lifetimes under an adaptive Type-II progressively censoring scheme. *Naval Research Logistics.* 56:687-698.

Table 1: Bias and MSEs of the MLEs for the PHC scheme under various censoring schemes.

(n, m)	Scheme	T=0.10		T=0.25		T=0.50	
		$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
(30,10)	1	0.373 (0.223)	0.425 (0.288)	0.272 (0.140)	0.293 (0.152)	0.252 (0.129)	0.262 (0.132)
	2	0.645 (0.741)	0.659 (0.814)	0.433 (0.328)	0.482 (0.401)	0.335 (0.207)	0.374 (0.237)
	3	0.385 (0.239)	0.435 (0.309)	0.279 (0.136)	0.302 (0.151)	0.248 (0.120)	0.268 (0.130)
(30,20)	1	0.395 (0.246)	0.448 (0.323)	0.255 (0.105)	0.289 (0.136)	0.200 (0.067)	0.225 (0.083)
	2	0.453 (0.342)	0.518 (0.439)	0.311 (0.160)	0.352 (0.202)	0.239 (0.092)	0.280 (0.127)
	3	0.410 (0.269)	0.454 (0.338)	0.289 (0.141)	0.324 (0.177)	0.230 (0.090)	0.262 (0.116)
(50,10)	1	0.303 (0.151)	0.330 (0.186)	0.238 (0.115)	0.257 (0.126)	0.235 (0.117)	0.254 (0.125)
	2	0.624 (0.701)	0.642 (0.795)	0.428 (0.328)	0.488 (0.414)	0.334 (0.205)	0.376 (0.245)
	3	0.301 (0.151)	0.326 (0.174)	0.251 (0.133)	0.260 (0.130)	0.241 (0.119)	0.258 (0.126)
(50,20)	1	0.294 (0.135)	0.342 (0.184)	0.198 (0.063)	0.228 (0.084)	0.175 (0.051)	0.200 (0.068)
	2	0.443 (0.336)	0.504 (0.431)	0.302 (0.147)	0.342 (0.195)	0.241 (0.097)	0.284 (0.130)
	3	0.306 (0.151)	0.357 (0.205)	0.217 (0.077)	0.255 (0.105)	0.183 (0.056)	0.211 (0.075)

Table 2: Bias and MSEs of the MLEs for the APC scheme under various censoring schemes.

(n, m)	Scheme	T=0.10		T=0.25		T=0.50	
		$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
(30,10)	1	0.267 (0.143)	0.310 (0.174)	0.260 (0.133)	0.258 (0.130)	0.251 (0.128)	0.260 (0.131)
	2	0.264 (0.149)	0.326 (0.208)	0.294 (0.184)	0.339 (0.222)	0.290 (0.184)	0.325 (0.204)
	3	0.273 (0.145)	0.310 (0.174)	0.272 (0.138)	0.292 (0.151)	0.242 (0.117)	0.259 (0.125)
(30,20)	1	0.189 (0.063)	0.224 (0.092)	0.188 (0.064)	0.215 (0.082)	0.190 (0.064)	0.216 (0.081)
	2	0.192 (0.065)	0.223 (0.087)	0.194 (0.066)	0.213 (0.084)	0.187 (0.062)	0.222 (0.088)
	3	0.192 (0.065)	0.222 (0.088)	0.192 (0.065)	0.221 (0.090)	0.188 (0.063)	0.216 (0.082)
(50,10)	1	0.270 (0.133)	0.279 (0.146)	0.236 (0.114)	0.254 (0.124)	0.235 (0.117)	0.254 (0.125)
	2	0.272 (0.157)	0.323 (0.229)	0.295 (0.191)	0.337 (0.234)	0.286 (0.169)	0.330 (0.211)
	3	0.269 (0.133)	0.284 (0.142)	0.246 (0.130)	0.252 (0.125)	0.241 (0.119)	0.258 (0.126)
(50,20)	1	0.183 (0.061)	0.221 (0.084)	0.185 (0.060)	0.215 (0.079)	0.173 (0.050)	0.197 (0.067)
	2	0.195 (0.068)	0.215 (0.084)	0.190 (0.064)	0.210 (0.080)	0.192 (0.065)	0.227 (0.091)
	3	0.190 (0.063)	0.218 (0.086)	0.192 (0.066)	0.222 (0.088)	0.179 (0.055)	0.207 (0.073)

Table 3: Bias and MSEs of the Bayes estimators for the APC scheme under various censoring schemes.

(n, m)	Scheme	T=0.10		T=0.25		T=0.50	
		$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
(30,10)	1	0.122 (0.025)	0.123 (0.023)	0.122 (0.024)	0.105 (0.017)	0.118 (0.023)	0.108 (0.018)
	2	0.118 (0.025)	0.127 (0.026)	0.130 (0.029)	0.130 (0.026)	0.129 (0.028)	0.125 (0.025)
	3	0.125 (0.026)	0.122 (0.023)	0.126 (0.025)	0.118 (0.021)	0.114 (0.022)	0.107 (0.018)
(30,20)	1	0.125 (0.026)	0.133 (0.030)	0.123 (0.025)	0.127 (0.026)	0.125 (0.026)	0.127 (0.026)
	2	0.127 (0.027)	0.133 (0.029)	0.127 (0.026)	0.125 (0.026)	0.123 (0.025)	0.131 (0.028)
	3	0.127 (0.027)	0.132 (0.030)	0.125 (0.026)	0.129 (0.028)	0.123 (0.025)	0.128 (0.027)
(50,10)	1	0.127 (0.025)	0.114 (0.020)	0.113 (0.021)	0.107 (0.018)	0.112 (0.021)	0.108 (0.019)
	2	0.122 (0.026)	0.124 (0.025)	0.129 (0.029)	0.128 (0.026)	0.128 (0.028)	0.126 (0.025)
	3	0.126 (0.025)	0.116 (0.020)	0.116 (0.023)	0.105 (0.018)	0.114 (0.022)	0.109 (0.018)
(50,20)	1	0.120 (0.024)	0.131 (0.028)	0.122 (0.025)	0.128 (0.026)	0.115 (0.021)	0.119 (0.023)
	2	0.128 (0.027)	0.127 (0.026)	0.124 (0.026)	0.124 (0.025)	0.125 (0.026)	0.133 (0.029)
	3	0.125 (0.025)	0.129 (0.028)	0.125 (0.026)	0.131 (0.028)	0.119 (0.023)	0.124 (0.025)

Table 4: The length of the different confidence intervals and the corresponding coverage probability (within parenthesis) the APC scheme under various censoring schemes.

(n, m)	Scheme	Method	T=0.10		T=0.25		T=0.50	
			$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
(30,10)	1	Approx.	1.390 (0.989)	1.569 (0.963)	1.342 (0.961)	1.495 (0.966)	1.336 (0.979)	1.456 (0.963)
		Bayes	0.914 (0.983)	0.953 (0.984)	0.893 (0.999)	0.932 (0.999)	0.885 (0.999)	0.916 (0.991)
		Boot-p	1.863 (0.945)	2.104 (0.945)	1.585 (0.966)	1.749 (0.979)	1.568 (0.969)	1.699 (0.977)
		Boot-t	1.401 (0.988)	1.568 (0.969)	1.274 (0.979)	1.393 (0.974)	1.305 (0.982)	1.409 (0.960)
(30,10)	2	Approx.	1.430 (0.988)	1.591 (0.961)	1.433 (0.981)	1.612 (0.963)	1.431 (0.982)	1.609 (0.967)
		Bayes	0.919 (0.997)	0.950 (0.989)	0.921 (0.998)	0.957 (0.999)	0.921 (0.998)	0.958 (0.997)
		Boot-p	1.910 (0.954)	2.128 (0.934)	1.897 (0.930)	2.124 (0.934)	1.863 (0.939)	2.079 (0.941)
		Boot-t	1.450 (0.987)	1.601 (0.961)	1.397 (0.980)	1.545 (0.964)	1.387 (0.986)	1.536 (0.972)
(30,10)	3	Approx.	1.391 (0.987)	1.576 (0.965)	1.367 (0.971)	1.539 (0.966)	1.334 (0.973)	1.467 (0.970)
		Bayes	0.913 (0.999)	0.954 (0.985)	0.903 (0.999)	0.944 (0.998)	0.887 (0.998)	0.921 (0.991)
		Boot-p	1.881 (0.945)	2.134 (0.941)	1.658 (0.951)	1.842 (0.964)	1.565 (0.974)	1.711 (0.982)
		Boot-t	1.405 (0.985)	1.581 (0.967)	1.304 (0.981)	1.439 (0.973)	1.286 (0.980)	1.397 (0.971)
(30,20)	1	Approx.	0.898 (0.943)	1.035 (0.938)	0.905 (0.947)	1.045 (0.959)	0.903 (0.947)	1.039 (0.955)
		Bayes	0.715 (0.977)	0.781 (0.974)	0.720 (0.985)	0.787 (0.990)	0.719 (0.980)	0.785 (0.987)
		Boot-p	1.064 (0.937)	1.246 (0.916)	1.043 (0.942)	1.211 (0.928)	1.004 (0.928)	1.155 (0.931)
		Boot-t	0.961 (0.950)	1.124 (0.941)	0.894 (0.949)	1.028 (0.959)	0.883 (0.946)	1.009 (0.956)
(30,20)	2	Approx.	0.902 (0.934)	1.035 (0.944)	0.907 (0.939)	1.042 (0.955)	0.900 (0.943)	1.038 (0.953)
		Bayes	0.717 (0.972)	0.781 (0.971)	0.720 (0.980)	0.785 (0.988)	0.716 (0.983)	0.783 (0.991)
		Boot-p	1.063 (0.931)	1.240 (0.919)	1.044 (0.936)	1.207 (0.930)	1.032 (0.943)	1.197 (0.920)
		Boot-t	0.964 (0.939)	1.124 (0.944)	0.896 (0.943)	1.027 (0.952)	0.888 (0.944)	1.021 (0.952)
(30,20)	3	Approx.	0.903 (0.936)	1.033 (0.935)	0.913 (0.955)	1.048 (0.951)	0.904 (0.950)	1.041 (0.954)
		Bayes	0.717 (0.969)	0.779 (0.976)	0.724 (0.988)	0.788 (0.988)	0.720 (0.982)	0.786 (0.987)
		Boot-p	1.067 (0.929)	1.239 (0.920)	1.049 (0.935)	1.210 (0.927)	1.035 (0.939)	1.199 (0.923)
		Boot-t	0.966 (0.938)	1.120 (0.935)	0.902 (0.960)	1.032 (0.950)	0.893 (0.947)	1.025 (0.955)
(50,10)	1	Approx.	1.341 (0.968)	1.498 (0.953)	1.314 (0.979)	1.432 (0.954)	1.317 (0.986)	1.415 (0.949)
		Bayes	0.894 (0.998)	0.932 (0.996)	0.880 (0.998)	0.921 (0.999)	0.880 (0.997)	0.918 (0.999)
		Boot-p	1.639 (0.952)	1.815 (0.971)	1.543 (0.967)	1.670 (0.981)	1.547 (0.969)	1.656 (0.977)
		Boot-t	1.284 (0.984)	1.410 (0.964)	1.276 (0.980)	1.378 (0.954)	1.303 (0.986)	1.394 (0.948)
(50,10)	2	Approx.	1.422 (0.981)	1.589 (0.963)	1.455 (0.986)	1.635 (0.959)	1.433 (0.990)	1.622 (0.967)
		Bayes	0.915 (0.996)	0.956 (0.997)	0.904 (0.996)	0.956 (0.997)	0.901 (0.998)	0.954 (0.999)
		Boot-p	1.871 (0.947)	2.082 (0.935)	1.923 (0.932)	2.155 (0.932)	1.862 (0.935)	2.096 (0.938)
		Boot-t	1.405 (0.983)	1.551 (0.966)	1.415 (0.985)	1.568 (0.963)	1.390 (0.990)	1.547 (0.973)
(50,10)	3	Approx.	1.343 (0.964)	1.513 (0.962)	1.332 (0.964)	1.444 (0.958)	1.320 (0.981)	1.431 (0.956)
		Bayes	0.908 (0.998)	0.949 (0.998)	0.886 (0.998)	0.915 (0.999)	0.881 (0.995)	0.910 (0.998)
		Boot-p	1.668 (0.949)	1.864 (0.964)	1.560 (0.959)	1.682 (0.981)	1.552 (0.970)	1.677 (0.983)
		Boot-t	1.291 (0.982)	1.430 (0.968)	1.283 (0.969)	1.378 (0.960)	1.306 (0.982)	1.410 (0.953)



Fully Stochastic Single-Period Mean-Variance Portfolio Optimization

R. Keykhaei¹, B. Panahbehagh²

¹Faculty of Mathematics and Computer of Khansar, Isfahan, Iran

²Department of Mathematics and computer science, Kharazmi University, Tehran, Iran

Abstract: The only sources of uncertainty in the standard Markowitz's single-period Mean-Variance portfolio selection problem are the future price of assets. In this paper we study the single-period Mean-Variance portfolio selection problem under general sources of uncertainty, which generalizes the Markowitz's model. Also we introduce another type of uncertainty which affects the quantity (number, weight, etc.) of each security contained in the portfolio. It is shown that under some general circumstances the set of optimal portfolios in the generalized model coincides with the standard Markowitz's model.

Keywords Single-period Mean-Variance portfolio optimization, Optimal portfolio, Uncertain exit time, Weight coefficient.

Mathematics Subject Classification (2010): 91G10 90C20.

1 introduction

The foundations of Modern Portfolio Theory (MPT) was established in the 1950's by [Markowitz \(1952, 1959\)](#). The aim of Markowitz's single-period Mean-Variance (M-V) portfolio selection problem is to maximize the expected return of a portfolio and minimize its variance as the measure of risk. M-V portfolio selection problems seek to compute *efficient* portfolios. A portfolio is efficient if with respect to its location in the M-V space, there is no obtainable portfolio with a lower variance without a lower expected return or a greater expected return without a greater variance. Analytical expression of the

¹R. Keykhaei: r.keykhaei@math.iut.ac.ir

M-V efficient portfolios in single-period case was derived by [Markowitz \(1952, 1959\)](#) and [Merton \(1972\)](#).

In fact a single period M-V problem is a static optimization problem which consider the future price of securities as the source of uncertainty. But, in the real world, there are other sources of uncertainty, such as uncertain time horizon, considered by other researchers. In this paper we are motivated by [Martellini and Urošević \(2005\)](#). They consider two types of uncertainty in their model. The first type is asset price uncertainty and the second type is called exit time risk, which derives from the uncertainty of real exit time of investors. In this paper we study the single-period M-V portfolio selection problem under general sources of uncertainty, which generalizes the model of [Markowitz \(1952\)](#) and covers the model of [Martellini and Urošević \(2005\)](#) model. In the following we introduce a third type of uncertainty which affects the quantity (number, weight, etc.) of each security contained in the portfolio, and combine it with uncertain exit time. In fact we assume that the quantity of each security can vary randomly during the time period of investment. It is shown that under some general circumstances the set of optimal portfolios in the generalized model coincides with the standard Markowitz's model.

The reminder of this paper is organized as follows. The basic notation, definitions and the formulation of standard single-period M-V portfolio selection problem are given in Section 2. In Section 3 we describe the notion of portfolio selection problem under general sources of uncertainty and reformulated it in a standard quadratic programming form. Some conditions are provided for which the sets of optimal portfolios are coincide in the standard and the general models. Two types of examples for the generalized model are given, distinctly, in Sections 4 and 5.

2 standard model

Consider an investor who enters in the market with initial wealth $X(0)$ and invests his/her wealth as $X(0) = \sum_{i=1}^n \phi_i S_i(0)$ between $n \geq 2$ risky securities, where $S_i(0)$ and $S_i(T)$ are, respectively, the value of the i th security at the beginning and the end of the period, and for each i , ϕ_i is the quantity of the i th security. At the end of the investment period ($t = T$) the wealth is $X(T) = \sum_{i=1}^n \phi_i S_i(T)$. Here ϕ_i 's remain fixed during the period.

In the classic Mean-Variance problem the return of each security is $r_i = \frac{S_i(T)}{S_i(0)}$ and the return of the portfolio is $r = \frac{X(T)}{X(0)}$. Then

$$r = \frac{\sum_{i=1}^n \phi_i S_i(T)}{X(0)} = \sum_{i=1}^n \frac{\phi_i S_i(0)}{X(0)} \frac{S_i(T)}{S_i(0)} = \sum_{i=1}^n x_i r_i,$$

in which

$$x_i = \frac{\phi_i S_i(0)}{X(0)}, \quad i = 1, \dots, n.$$

Clearly $\sum_{i=1}^n x_i = 1$. Indeed for each i , x_i is the the assigned weight allocated to the i th security in the portfolio. We denote the portfolio by $\mathbf{X} := (x_1, \dots, x_n)' \in \mathbb{R}^n$. Also let $R = (r_1, \dots, r_n)'$ be the random vector of returns with covariance matrix $\mathbf{V} = (\sigma_{ij})_{n \times n}$ and mean vector $\bar{\mathbf{R}} = (\bar{r}_1, \dots, \bar{r}_n)'$. Here we assume that \mathbf{V} is positive definite. Obviously $\mathbb{E}(r) = \bar{\mathbf{R}}' \mathbf{X}$ and $\text{Var}(r) = \mathbf{X}' \mathbf{V} \mathbf{X}$. The M-V problem for a desired return μ^* is:

$$\textbf{Problem 1:} \quad \min_{\mathbf{X}} \frac{1}{2} \mathbf{X}' \mathbf{V} \mathbf{X}, \quad \text{s.t.} \quad \{\bar{\mathbf{R}}' \mathbf{X} = \mu^*, \mathbf{1}' \mathbf{X} = 1\},$$

where $\mathbf{1}$ is the n -column vector of ones.

Theorem 2.1. *Problem 1 have the unique optimal solution*

$$\mathbf{X} = \frac{C - \mu^* B}{AC - B^2} \mathbf{V}^{-1} \mathbf{1} + \frac{\mu^* A - B}{AC - B^2} \mathbf{V}^{-1} \bar{\mathbf{R}},$$

where

$$A := \mathbf{1}' \mathbf{V}^{-1} \mathbf{1}, \quad B := \mathbf{1}' \mathbf{V}^{-1} \bar{\mathbf{R}} = \bar{\mathbf{R}}' \mathbf{V}^{-1} \mathbf{1}, \quad C := \bar{\mathbf{R}}' \mathbf{V}^{-1} \bar{\mathbf{R}}.$$

Proof. See chapter 4 of [Ingersoll \(1987\)](#). □

3 Fully Stochastic Portfolio Selection

Let the wealth of the investor is affected by some sources of uncertainty denoted by $\Theta = (\theta_1, \dots, \theta_m)' \in \mathbb{R}^m$. Then $r^\Theta = \frac{X(\Theta)}{X(0)}$ is the return of the investment, where $\mathbf{X}(\Theta)$ is the wealth of the investor at the exit time. Actually, in the standard case $\Theta = (S_1, \dots, S_n)$, where S_i denotes the price of the i th security, for $i = 1, \dots, n$. The fully stochastic single-period M-V portfolio selection problem for a desired return μ is:

$$\textbf{Problem 2:} \quad \min_{\mathbf{X}} \text{Var}(r^\Theta) \quad \text{s.t.} \quad \{\mathbb{E}(r^\Theta) = \mu, \quad \mathbf{1}' \mathbf{X} = 1\}.$$

In the following, for comparing with the standard case, we fix the S_i 's for $i = 1, \dots, n$, as the trivial sources of uncertainty and consider Θ for the other sources. The aim is to describe the return of the portfolio by the convex combination

$$r^\Theta = \mathbf{X}'\mathbf{R}^\Theta,$$

where $\mathbf{R}^\Theta = (r_1^\Theta, \dots, r_n^\Theta)'$ is the random vector of returns $r_i^\Theta := r_i^\Theta(S_i, \Theta)$ with covariance matrix Σ and mean vector $\bar{\mathbf{R}}^\Theta = (\bar{r}_1^\Theta, \dots, \bar{r}_n^\Theta)'$. Then Problem 1 can be reformulated as a standard quadratic program as follows:

$$\textbf{Problem 3:} \quad \min_{\mathbf{X}} \frac{1}{2}\mathbf{X}'\Sigma\mathbf{X} \quad \text{s.t.} \quad \{\mathbf{X}'\bar{\mathbf{R}}^\Theta = \mu, \quad \mathbf{X}'\mathbf{1} = 1\}.$$

In the following we try to provide some conditions for which the generalized and the standard problems have the same set of optimal portfolios when desired mean return varies over all obtainable mean returns. This means that the other sources of uncertainty does not effect on our optimal choice. Let $g, h : \mathbb{R}^m \rightarrow \mathbb{R}$ are real-valued functions and

$$\mathbb{E}(r_i^\Theta | \Theta = \theta) = \bar{r}_i g(\theta), \quad (3.1)$$

$$\text{Cov}(r_i^\Theta, r_j^\Theta | \Theta = \theta) = \sigma_{ij} h(\theta). \quad (3.2)$$

Then

$$\mathbb{E}(r_i^\Theta) = \int_{\mathbb{R}^m} \mathbb{E}(r_i^\Theta | \Theta = \theta) dF(\theta) = \int_{\mathbb{R}^m} \bar{r}_i g(\theta) dF(\theta) = \bar{r}_i \mathbb{E}(g(\Theta)),$$

where F denotes the distribute function corresponding to Θ . Using the equation

$$\text{Cov}(r_i^\Theta, r_j^\Theta) = \mathbb{E}(\text{Cov}(r_i^\Theta, r_j^\Theta | \Theta)) + \text{Cov}(\mathbb{E}(r_i^\Theta | \Theta), \mathbb{E}(r_j^\Theta | \Theta))$$

we can calculate the covariation between r_i^Θ and r_j^Θ . We can see

$$\begin{aligned} \mathbb{E}(\text{Cov}(r_i^\Theta, r_j^\Theta | \Theta)) &= \int_{\mathbb{R}^m} \text{Cov}(r_i^\Theta, r_j^\Theta | \Theta = \theta) dF(\theta) \\ &= \int_{\mathbb{R}^m} \sigma_{ij} h(\theta) dF(\theta) = \sigma_{ij} \mathbb{E}(h(\Theta)) \end{aligned}$$

and

$$\text{Cov}(\mathbb{E}(r_i^\Theta | \Theta), \mathbb{E}(r_j^\Theta | \Theta)) = \mathbb{E}(\mathbb{E}(r_i^\Theta | \Theta)\mathbb{E}(r_j^\Theta | \Theta)) - \mathbb{E}(\mathbb{E}(r_i^\Theta | \Theta))\mathbb{E}(\mathbb{E}(r_j^\Theta | \Theta))$$

$$\begin{aligned}
&= \int_{\mathbb{R}^m} \mathbb{E}(r_i^\Theta | \Theta = \theta) \mathbb{E}(r_j^\Theta | \Theta = \theta) dF(\theta) - \bar{r}_i \bar{r}_j \mathbb{E}^2(g(\Theta)) \\
&= \int_{\mathbb{R}^m} \bar{r}_i \bar{r}_j g^2(\theta) dF(\theta) - \bar{r}_i \bar{r}_j \mathbb{E}^2(g(\Theta)) \\
&= \bar{r}_i \bar{r}_j \text{Var}(g(\Theta)).
\end{aligned}$$

Therefore $\text{Cov}(r_i^\Theta, r_j^\Theta) = \sigma_{ij} \mathbb{E}(h(\Theta)) + \bar{r}_i \bar{r}_j \text{Var}(g(\Theta))$. Then

$$\mathbf{\Sigma} = \mathbb{E}(h(\Theta)) \mathbf{V} + \text{Var}(g(\Theta)) \mathbf{E}, \quad (3.3)$$

where $\mathbf{E} = (e_{ij})_{n \times n}$ and $e_{ij} = \bar{r}_i \bar{r}_j$ for $i, j = 1, \dots, n$. If $\mathbf{X}' \bar{\mathbf{R}} = \mu^*$ then we have

$$\mu := \mathbb{E}(r^\Theta) = \mathbf{X}' \bar{\mathbf{R}}^\Theta = \mathbb{E}(g(\Theta)) \mathbf{X}' \bar{\mathbf{R}} = \mathbb{E}(g(\Theta)) \mu^*,$$

$$\text{Var}(r^\Theta) = \mathbf{X}' \mathbf{\Sigma} \mathbf{X} = \mathbb{E}(h(\Theta)) \mathbf{X}' \mathbf{V} \mathbf{X} + \text{Var}(g(\Theta)) \mathbf{X}' \mathbf{E} \mathbf{X}.$$

Lemma 3.1. *The matrix $\mathbf{\Sigma}$ is positive definite if $\mathbb{E}(h(\Theta)) > 0$.*

Proof. Let $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{X} \neq 0$. Then

$$\mathbf{X}' \mathbf{\Sigma} \mathbf{X} = \mathbb{E}(h(\Theta)) \mathbf{X}' \mathbf{V} \mathbf{X} + \text{Var}(g(\Theta)) (\bar{\mathbf{R}}' \mathbf{X})^2 > 0,$$

where the last inequality holds since \mathbf{V} is positive definite and $\text{Var}(g(\Theta))$ is non-negative. \square

Theorem 3.2. *If $\mathbb{E}(h(\Theta)) > 0$, then problems 1 and 3 have the same unique optimal solution for which $\mu^* = \frac{\mu}{\mathbb{E}(g(\Theta))}$.*

Proof. Using equation (3.3) the Lagrangian function corresponding to problem 3 is

$$L(\mathbf{X}, \lambda_1, \lambda_2) := \frac{\mathbb{E}(h(\Theta))}{2} \mathbf{X}' \mathbf{V} \mathbf{X} + \frac{\text{Var}(g(\Theta))}{2} \mathbf{X}' \mathbf{E} \mathbf{X} + \lambda_1 (\mathbf{X}' \bar{\mathbf{R}}^\Theta - \mu) + \lambda_2 (\mathbf{X}' \mathbf{1} - 1).$$

Considering the first optimality condition we should have

$$\frac{\partial L}{\partial \mathbf{X}} = \mathbb{E}(h(\Theta)) \mathbf{V} \mathbf{X} + \text{Var}(g(\Theta)) \mathbf{E} \mathbf{X} + \lambda_1 \bar{\mathbf{R}}^\Theta + \lambda_2 \mathbf{1} = \mathbf{0}, \quad (3.4)$$

$$\frac{\partial L}{\partial \lambda_1} = \mathbf{X}' \bar{\mathbf{R}}^\Theta - \mu = 0, \quad \frac{\partial L}{\partial \lambda_2} = \mathbf{X}' \mathbf{1} - 1 = 0. \quad (3.5)$$

From (3.5) we have $\bar{\mathbf{R}}' \mathbf{X} = \mu^*$ and then $\mathbb{E}(g(\Theta)) \mathbf{E} \mathbf{X} = \mu^* \bar{\mathbf{R}}$. So equation (3.4) can be rewritten as

$$\mathbb{E}(h(\Theta)) \mathbf{V} \mathbf{X} + \text{Var}(g(\Theta)) \frac{\mu}{\mathbb{E}(g(\Theta))} \bar{\mathbf{R}} + \lambda_1 \mathbb{E}(g(\Theta)) \bar{\mathbf{R}} + \lambda_2 \mathbf{1} = \mathbf{0}.$$

Then

$$\mathbf{X} = \alpha \mathbf{V}^{-1} \mathbf{1} + \beta \mathbf{V}^{-1} \bar{\mathbf{R}}, \quad (3.6)$$

where $\alpha = -\frac{\lambda_2}{\mathbb{E}(h(\Theta))}$ and $\beta = -\left(\frac{\text{Var}(g(\Theta))\mu}{\mathbb{E}(h(\Theta))\mathbb{E}(g(\Theta))} + \frac{\lambda_1 \mathbb{E}(g(\Theta))}{\mathbb{E}(h(\Theta))}\right)$. Replacing \mathbf{X} in equations (3.5) yields

$$\mathbf{1} = \mathbf{1}'\mathbf{X} = \alpha \mathbf{1}'\mathbf{V}^{-1} \mathbf{1} + \beta \mathbf{1}'\mathbf{V}^{-1} \bar{\mathbf{R}} = \alpha A + \beta B \quad (3.7)$$

$$\mu^* = \frac{\mu}{\mathbb{E}(g(\Theta))} = \bar{\mathbf{R}}'\mathbf{X} = \alpha \bar{\mathbf{R}}'\mathbf{V}^{-1} \mathbf{1} + \beta \bar{\mathbf{R}}'\mathbf{V}^{-1} \bar{\mathbf{R}} = \alpha B + \beta C \quad (3.8)$$

After solving the system (3.7)-(3.8) we have

$$\alpha = \frac{C - \mu^* B}{AC - B^2}, \quad \beta = \frac{\mu^* A - B}{AC - B^2} \quad (3.9)$$

Now the claim follows from equations (3.6) and (3.9) and theorem 2.1. The uniqueness of the solution follows from positive definiteness of Σ . \square

Theorem 3.2 guaranties that the set of optimal portfolios of the standard case and the generalized case are the same, under conditions (3.1) and (3.2). In the following, two examples of the generalized problem are presented. The first one is taken from [Martellini and Urošević \(2005\)](#).

4 Uncertain Exit Time

[Martellini and Urošević \(2005\)](#) first propose the concept of exit time risk and show that the set of M-V optimal portfolios in the case where the uncertain exit time τ is independent of the portfolio performance and the asset prices follow a random walk, coincides with the standard case. Although, when the exit time is dependent on asset returns, it is possible that an optimal portfolio in the standard case is not optimal in the generalized case and vice versa. They assume that the exit time is the same for all assets. In fact they set $\Theta = \tau$ and $r^\tau = \mathbf{X}'\mathbf{R}^\tau$, where

$$r_i^\tau := r_i^\tau(S_i, \tau) = \frac{S_i(\tau)}{S_i(0)}, \quad i = 1, \dots, n$$

and $S_i(\tau)$ is the price of the i th asset. When the exit time is independent and the asset prices follow a random walk we have $\mathbb{E}(r_i^\tau | \tau = t) = \mathbb{E}(r_i(t)) = \bar{r}_i t$ and $\text{Cov}(r_i^\tau, r_j^\tau | \tau = t) = \text{Cov}(r_i(t), r_j(t)) = \sigma_{ij} t$. Considering the functions $g, h : \mathbb{R} \rightarrow \mathbb{R}$ where $g(x) = h(x) = x$ in (3.1) and (3.2) we have the following theorem.

Theorem 4.1. (Proposition 4 of *Martellini and Urošević (2005)*) If the exit time τ is independent of portfolio performance and the asset prices follow a random walk, problems 1 and 3 have the same unique optimal solution for which $\mu^* := \frac{\mu}{\mathbb{E}(\tau)}$.

5 Stochastic Weights

In this model, following Section 1, we assume that ϕ_i 's are non-constant and can vary randomly during the time period of investment, for $i = 1, \dots, n$. In fact it is possible that $\phi_i(0) \neq \phi_i(T)$ for some i , where for each i , $\phi_i(0)$ and $\phi_i(T)$ are the quantity of the i th security at the beginning and at the end of the period, respectively. Thus the value of portfolio at the beginning and at the end of the period is $X(0) = \sum_{i=1}^n \phi_i(0)S_i(0)$ and $X(T) = \sum_{i=1}^n \phi_i(T)S_i(T)$, respectively.

Definition 5.1. The weight coefficient of the i th security is

$$\Lambda_i := \frac{\phi_i(T)}{\phi_i(0)}, \quad i = 1, \dots, n.$$

Example 5.2. Assume that an event occurs according to a Poisson process $\{N(t) : t \geq 0\}$ with rate ν , and for any arrival, the investor loses half amount of the i th security, where i is fixed. Then at the end of the period ($t = T$), with $N(T) = n$ for $n = 0, 1, 2, \dots$, we have

$$\phi_i(T) = \phi_i(0)/2^n.$$

Thus Λ_i is a discrete random variable such that

$$P(\Lambda_i = 1/2^n) = P(N(T) = n) = e^{-\nu T} \frac{(\nu T)^n}{n!}, \quad n = 0, 1, 2, \dots$$

Now by applying the notion of weight coefficients we have

$$r^\Theta = \frac{X(T)}{X(0)} = \frac{\sum_{i=1}^n \phi_i(T)S_i(T)}{X(0)} = \sum_{i=1}^n \frac{\phi_i(0)S_i(0)}{X(0)} \frac{\phi_i(T)}{\phi_i(0)} \frac{S_i(T)}{S_i(0)} = \sum_{i=1}^n x_i \Lambda_i r_i$$

for $x_i = \phi_i(0)S_i(0)/X(0)$ as the initial weight allocated to the i th security.

Definition 5.3. For each $i = 1, \dots, n$, we define $r_i^\Theta := \Lambda_i r_i$ as the total-return of the i th security, where $\Theta = (\Lambda_1, \dots, \Lambda_n)$. So $r^\Theta = \sum_{i=1}^n x_i r_i^\Theta = \mathbf{X}' \mathbf{R}^\Theta$.

In the following we assume that the all securities have the same weight coefficient, that is $\Lambda_i = \Lambda$ for $i = 1, \dots, n$, and independent of the security prices. We can see

$$\mathbb{E}(r_i^\ominus | \Theta = (\lambda, \dots, \lambda)) = \mathbb{E}(\lambda r_i | \Theta = (\lambda, \dots, \lambda)) = \lambda \bar{r}_i$$

and

$$\mathbb{E}(r_i^\ominus r_j^\ominus | \Theta = (\lambda, \dots, \lambda)) = \mathbb{E}(\lambda^2 r_i r_j | \Theta = (\lambda, \dots, \lambda)) = \lambda^2 \mathbb{E}(r_i r_j).$$

Therefore

$$\begin{aligned} \text{Cov}(r_i^\ominus, r_j^\ominus | \Theta = (\lambda, \dots, \lambda)) &= \mathbb{E}(r_i^\ominus r_j^\ominus | \Theta = (\lambda, \dots, \lambda)) \\ &\quad - \mathbb{E}(r_i^\ominus | \Theta = (\lambda, \dots, \lambda)) \mathbb{E}(r_j^\ominus | \Theta = (\lambda, \dots, \lambda)) \\ &= \lambda^2 \mathbb{E}(r_i r_j) - \lambda^2 \mathbb{E}(r_i) \mathbb{E}(r_j) \\ &= \lambda^2 \sigma_{ij}. \end{aligned}$$

Using the functions $g, h : \mathbb{R}^n \rightarrow \mathbb{R}$ where $g(x_1, \dots, x_n) = x_1$ and $h(x_1, \dots, x_n) = x_1^2$ in (3.1) and (3.2) we have the following theorem.

Theorem 5.4. *If all securities have the same weight coefficient Λ and independent of security returns, problems 1 and 3 have the same unique optimal solution for which $\mu^* := \frac{\mu}{\mathbb{E}(\Lambda)}$.*

References

- Ingersoll, J. (1987). *Theory of financial decision making*. Rowman and Littlefield, Totowa, NJ.
- Markowitz, H. (1952). Portfolio selection, *Journal of finance*, **7**, 77–91.
- Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investment*. John Wiley and Sons, New York.
- Martellini, L. and Urosevic, B. (2005). Static mean variance analysis with uncertain time-horizon. *Management Science* **52** (6), 955–964.
- Merton, R.C. (1972). An analytic derivation of the efficient portfolio. *The Journal of Financial and Quantitative Analysis*, **7**(4), 1852–1872.



Asymptotic Coverage Levels of Confidence Intervals for prediction in Functional Linear Regression

Omid Khademnoe¹, S. Mohammad E. Hosseini-Nasab²

^{1,2}Department of Statistics, Shahid Beheshti University, Tehran, Iran

Abstract: Functional linear regression models have drawn much attention recently. In this paper, we focus on a functional linear model with functional covariate and scalar response. We use the stochastic expansions of eigenvalues and eigenfunctions estimators of the covariance operator to construct confidence intervals for prediction in this model. Then, we obtain asymptotic coverage levels for these confidence intervals. Finally, we conduct a simulation study to illustrate the numerical performance of confidence intervals, under various parameter settings.

Keywords: Central limit theorem, Cross-validation, Functional linear regression model, Functional principal component analysis, Prediction.

Mathematics Subject Classification (2010): 62J99 62G99.

1 Introduction

In recent years, functional data analysis (FDA) has become more common because advances in technology facilitate collecting and storing the data that are essentially in the form of curves. The theoretical and methodological developments for FDA are reviewed by [Ramsay and Silverman \(2005\)](#).

Functional linear regression analysis is one of the most widely used techniques for modelling a variable of interest based on a set of related predictor variables. Here, we consider a functional linear regression model in which the response is scalar and the covariate of interest is functional.

When estimating the slope function, dimension reduction of the functional predictor and the slope function can be taken into account; see for example, [Cardot et al. \(1999\)](#) and [Yao et al. \(2005\)](#). Methods based on functional principal component analysis (FPCA) are popular and widely used by

¹Omid Khademnoe : o_khademnoe@sbu.ac.ir

researchers, since it allows finite dimensional analysis of a problem that is intrinsically infinite dimensional.

Cai and Hall (2006) studied the rate of convergence for prediction in the functional linear regression model. The rate of convergence to estimate the regression weight function in a functional linear regression model in which the predictor as well as the weight function are smoothed and periodic functions was investigated by Li and Hsing (2007). James et al. (2009) proposed an approach to obtain both interpretable, flexible and accurate estimate of the slope function. Cardot et al. (2007) gave a CLT for the functional linear regression model with scalar response in weak topology (or pointwise). However, this work ignores some random terms when obtaining the asymptotic distribution of prediction so that the impact of these terms on the weak convergence of the prediction has not been well understood. Taking these terms into account, we construct the confidence interval for prediction. Then, we obtain its asymptotic coverage level.

The paper is organized as follows. In Section 2, we first introduce functional linear models and then describe FPCA method for estimation of the slope function. Section 3 contains main results of our work, where we construct a confidence interval for prediction and obtain its asymptotic coverage level. In Section 4, we carry out a simulation study to illustrate the performance of the confidence interval for prediction under various situations.

2 Functional linear model

In functional linear regression models with scalar response and functional covariate, we assume that X is a square integrable random function defined on $\mathcal{I} = [0, T]$, and the real-valued response Y is generated by the model

$$Y = a + \int_{\mathcal{I}} \rho(u) X(u) du + \varepsilon, \quad (2.1)$$

where a and ρ denote the intercept and slope function in the model, respectively. We assume that a is a constant, ρ is a square integrable function on \mathcal{I} , the error ε is a real-valued variable such that $E[\varepsilon] = 0$, $E[\varepsilon^2] = \sigma^2 < \infty$ and also the error is independent of the X . Without loss of generality, we can drop the intercept a in (2.1).

A general review of the slope estimation based on the use of basis expansions can be found in [Ramsay and Silverman \(2005\)](#). We use the standard approach for estimating the slope function in the functional linear regression that is explicitly based on FPCA, allowing finite-dimensional analysis of the problem. For this purpose, we assume the covariance operator of X is positive definite, allowing a spectral decomposition in terms of strictly positive eigenvalues λ_j ,

$$C(u, v) = \text{cov}\{X(u), X(v)\} = \sum_{j=1}^{\infty} \lambda_j e_j(u) e_j(v), \quad u, v \in \mathcal{I}, \quad (2.2)$$

where the pairs (λ_j, e_j) are, respectively, eigenvalues and eigenfunctions of the integral operator with kernel C . To reduce the notational load, we shall also denote the operator by C . We assume that the eigenvalues $\lambda_1, \lambda_2, \dots$ and eigenfunctions e_1, e_2, \dots form a strictly decreasing sequence and an orthonormal basis for the space of all square-integrable functions on \mathcal{I} , denoted by $\mathcal{L}^2(\mathcal{I})$, respectively.

Assume we observe the independent data $(X_1, Y_1), \dots, (X_n, Y_n)$, each distributed as (X, Y) . The estimator of $C(u, v)$ and of its corresponding spectral decomposition are

$$\begin{aligned} \hat{C}(u, v) &= \frac{1}{n} \sum_{i=1}^n \{X_i(u) - \bar{X}(u)\} \{X_i(v) - \bar{X}(v)\} \\ &= \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{e}_j(u) \hat{e}_j(v), \quad u, v \in \mathcal{I}, \end{aligned}$$

where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$. Similarly, the eigenvalues and eigenfunctions of the integral operator \hat{C} with kernel $\hat{C}(u, v)$ are denoted by $\hat{\lambda}_j$ and \hat{e}_j , respectively. The function ρ can be expanded in terms of its Fourier series, as $\sum_{j \geq 1} \rho_j e_j$, where $\rho_j = \int_{\mathcal{I}} \rho(u) e_j(u) du$. We define the function $g(u) = \int_{\mathcal{I}} C(u, v) \rho(v) dv$ and denote the j th Fourier coefficient of g by $g_j = \int_{\mathcal{I}} g(u) e_j(u) du$.

Let $\hat{g}(u) = \frac{1}{n} \sum_{i=1}^n (X_i(u) - \bar{X}(u))(Y_i - \bar{Y})$ be an estimator of g ([Cai and Hall \(2006\)](#)). Noting that $g_j = \lambda_j \rho_j$, we take $\hat{\rho}_j = \hat{\lambda}_j^{-1} \hat{g}_j$, for $1 \leq j \leq m$, where $\hat{g}_j = \int_{\mathcal{I}} \hat{g}(u) \hat{e}_j(u) du$. Then, we estimate ρ by $\hat{\rho} = \sum_{j=1}^m \hat{\rho}_j \hat{e}_j$, where the cut-off point m lies in the range $1 \leq m \leq n$. Hence, for a fixed future observation x , the estimator of $E[Y | X = x]$ is obtained by $\hat{y} = \int_{\mathcal{I}} \hat{\rho}(u) x(u) du$.

3 Assumptions and main results

Let Z have the distribution of a generic $X - \mu$, where μ is the mean function of X . The Karhunen-Loève expansion of Z is written by $Z = \sum_{j \geq 1} U_j e_j$, where $U_j = \int_{\mathcal{I}} Z(u) e_j(u) du$ is the j th principal component, the U_j have zero means, and $E[U_j U_k] = \delta_{jk} \lambda_j$, where δ_{jk} is the Kronecker delta. We assume that

$$E[\|X\|^4] = E\left[\left(\int_{\mathcal{I}} X^2\right)^2\right] < \infty, \quad (3.1)$$

and, for each $j \geq 1$, $E|U_j|^4 \leq \text{constant} \cdot \lambda_j^2$, where the constant does not depend on j . We also suppose that, there exists a convex positive function λ such that at least for large j ,

$$\lambda_j = \lambda(j). \quad (3.2)$$

Define the finite rank operators C^\dagger and \hat{C}^\dagger with the same eigenfunctions as C and \hat{C} and associated eigenvalues λ_j^{-1} and $\hat{\lambda}_j^{-1}$, respectively, i.e.

$$C^\dagger = \sum_{j=1}^m \lambda_j^{-1} e_j \otimes e_j, \quad \hat{C}^\dagger = \sum_{j=1}^m \hat{\lambda}_j^{-1} \hat{e}_j \otimes \hat{e}_j,$$

where $e_j \otimes e_j(h) = \langle h, e_j \rangle e_j$, for every $h \in \mathcal{L}^2(\mathcal{I})$, and $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathcal{L}^2(\mathcal{I})$. Regarding the above notations, \hat{g} can be written as

$$\hat{g} = \hat{C}\rho + V_n, \quad (3.3)$$

where $V_n = n^{-1} \sum_{i=1}^n (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})$.

We denote the projectors onto the eigenspace corresponding to the first m eigenvalues of C and \hat{C} by Π_m and $\hat{\Pi}_m$, respectively. Thus, we have

$$\hat{\rho} - \rho = W_n + R_n + B_n, \quad (3.4)$$

where

$$W_n = (\hat{C}^\dagger - C^\dagger)V_n, \quad R_n = C^\dagger V_n, \quad B_n = \hat{\Pi}_m \rho - \rho.$$

We show that the inner products of the two terms W_n , B_n and the future observation x converge to zero in probability if correctly normalized. Moreover the inner product of R_n and x tends to a normal

distribution. We shall take all terms on the right-hand side of equation (3.4) into account to derive the asymptotic distribution of $\sqrt{n}\langle\hat{\rho} - \rho, x\rangle$, rather than the first two terms used by Cardot et al. (2007).

Proposition 3.1. *Suppose that*

$$\langle x, e_j \rangle^2 \lambda_j^{-1} < \infty, \text{ for all } j \geq 1, \quad (3.5)$$

$$n^{-1} t_{n,x}^{-2} m^3 (\log m)^2 \rightarrow 0 \quad \text{and} \quad n^{-1} t_{n,x}^{-1} m^{2.5} (\log m)^2 \rightarrow 0, \quad (3.6)$$

where $t_{n,x}^2 = \sum_{j=1}^m \lambda_j^{-1} \langle x, e_j \rangle$. If assumptions (3.1) and (3.2) hold, then $\sqrt{\frac{n}{t_{n,x}}} \langle W_n, x \rangle \rightarrow 0$ in probability, as $n \rightarrow \infty$.

Proof. Note that $\sqrt{n} \langle W_n, x \rangle = \sqrt{n} \sum_{j=1}^m T_j$, where

$$\begin{aligned} T_j &= \hat{\lambda}_j^{-1} \langle \hat{e}_j, x \rangle n^{-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon}) \langle \hat{e}_j, X_i - \bar{X} \rangle \\ &\quad - \lambda_j^{-1} \langle e_j, x \rangle n^{-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon}) \langle e_j, X_i - \bar{X} \rangle. \end{aligned} \quad (3.7)$$

Using the stochastic expansions of eigenvalues and eigenfunctions estimators of the covariance operator C appeared in Hall and Hosseini-Nasab (2006, 2009), it can be proved that

$$\frac{n}{t_{n,x}^2} E[\langle W_n, x \rangle^2] = \frac{n}{t_{n,x}^2} \sum_{j=1}^m E[T_j^2] + o(1). \quad (3.8)$$

The proof of Proposition 3.1 will be complete if we show that the term on the right-hand side of (3.8) tends to zero. Using (3.2), (3.5) and (3.6) and applying Cauchy-Schwartz and Jensen's inequalities, the desired result is proved. \square

Note that the two conditions appeared in (3.6) are milder than the two following conditions used by Cardot et al. (2007):

$$\sum_{k=1}^{\infty} |\langle \rho, e_j \rangle| < \infty, \quad (3.9)$$

$$n^{-0.5} t_{n,x}^{-1} m^3 (\log m)^2 \rightarrow 0, \quad \text{and} \quad n^{-0.5} m^{2.5} (\log m)^2 \rightarrow 0. \quad (3.10)$$

Proposition 3.2. *If assumption (3.1) holds, then $\sqrt{n} \langle R_n, x \rangle \rightarrow N(0, \sigma^2 t_{n,x}^2)$ in distribution, as $n \rightarrow \infty$.*

Proof. We have

$$\begin{aligned} \langle R_n, x \rangle &= \left\langle \left(\sum_{j=1}^m \lambda_j^{-1} e_j \otimes e_j \right) \left(n^{-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})(X_i - \bar{X}) \right), x \right\rangle \\ &= n^{-1} \sum_{i=1}^n S_i + o_p(n^{-1/2}), \end{aligned}$$

where $S_i = \varepsilon_i \sum_{j=1}^m \lambda_j^{-1} \langle e_j, X_i - \mu \rangle \langle e_j, x \rangle$. The variables S_i are independent and identically distributed. Therefore, by applying the Central Limit Theorem to this sequence, we have

$$n^{-1} \sum_{i=1}^n S_i \rightarrow N(0, n^{-1} \sigma^2 t_{n,x}^2), \text{ in distribution.}$$

□

By the above results, the nominal $(1 - \alpha)$ -level asymptotic confidence intervals for prediction in the functional linear model can be constructed as follows:

$$CI_{asy}(\alpha) = (\hat{y} - n^{-1/2} z_{\alpha/2} \hat{\sigma} \hat{t}_{n,x}, \hat{y} + n^{-1/2} z_{\alpha/2} \hat{\sigma} \hat{t}_{n,x}),$$

where $\hat{\sigma}$ and $\hat{t}_{n,x}$ are estimated values of σ and $t_{n,x}$, respectively, and z_α is the α -level quantile of the standard normal distribution.

In Remark 5 of Cardot et al. (2007), it is pointed out that convergence of the inner product of the bias term (B_n) and a fixed predictor to zero requires very restrictive conditions either on ρ or on the eigenvalues λ_j . Here, we are going to derive the same results under some assumptions milder than those used by Cardot et al. (2007).

Theorem 3.3. *Assume that for all $j \geq 1$,*

$$(|\langle \rho, e_j \rangle| j^\beta) < \infty, \text{ for some } \beta > 1, \quad (3.11)$$

$$(|\langle x, e_j \rangle| j^\gamma) < \infty, \text{ for some } \gamma > 1 \text{ and } \beta + \gamma \geq 3.5, \quad (3.12)$$

$$\sqrt{nm}^{-2.5} t_{n,x}^{-1} \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (3.13)$$

Then, $P(E[Y|X = x] \in CI_{asy}(\alpha)) = 1 - \alpha + E_n$ such that $E_n \rightarrow 0$ as $n \rightarrow \infty$.

Proof. Under assumptions (3.1), (3.2) and (3.11)-(3.13), we have $\sqrt{n} \langle B_n, x \rangle \rightarrow 0$ in probability, as $n \rightarrow \infty$. Since $\hat{t}_{n,x} \rightarrow t_{n,x}$ and $\hat{\sigma} \rightarrow \sigma$, in probability, we have $\frac{\sqrt{n}}{\hat{t}_{n,x} \hat{\sigma}} \langle (\hat{\rho} - \rho), x \rangle \rightarrow N(0, 1)$, in distribution, as $n \rightarrow \infty$. □

Table 1: Coverage levels of confidence intervals with nominal coverage level $1 - \alpha = 0.95$. The notations E, U and N denote the centred exponential, uniform and normal distribution, respectively.

n	α	β	γ	Continuous			Discrete			
				Distribution of U_j			Distribution of U_j			
				E	U	N	E	U	N	
50	1.5	1	3	0.89(0.29)	0.89(0.28)	0.92(0.30)	0.92(0.29)	0.92(0.28)	0.92(0.29)	
			2	3	0.87(0.28)	0.88(0.27)	0.91(0.28)	0.92(0.29)	0.92(0.28)	0.93(0.28)
			2	2	0.83(0.30)	0.85(0.29)	0.91(0.30)	0.92(0.31)	0.91(0.30)	0.95(0.31)
		3	2	0.87(0.30)	0.87(0.29)	0.91(0.30)	0.92(0.30)	0.93(0.30)	0.94(0.30)	
			3	1	0.81(0.36)	0.83(0.35)	0.92(0.38)	0.88(0.40)	0.88(0.39)	0.96(0.40)
			2	3	0.93(0.14)	0.93(0.14)	0.93(0.15)	0.94(0.14)	0.93(0.14)	0.94(0.14)
200	1.5	1	3	0.92(0.14)	0.92(0.14)	0.94(0.14)	0.94(0.14)	0.94(0.14)	0.95(0.14)	
			2	2	0.89(0.15)	0.90(0.15)	0.95(0.15)	0.93(0.15)	0.93(0.15)	0.96(0.16)
			3	2	0.92(0.15)	0.92(0.15)	0.95(0.15)	0.94(0.15)	0.94(0.15)	0.96(0.15)
		3	2	0.92(0.14)	0.92(0.14)	0.94(0.14)	0.94(0.14)	0.94(0.14)	0.95(0.14)	
			3	1	0.86(0.19)	0.86(0.19)	0.95(0.20)	0.91(0.20)	0.91(0.21)	0.97(0.21)
			2	3	0.93(0.14)	0.93(0.14)	0.93(0.15)	0.94(0.14)	0.93(0.14)	0.94(0.14)

For the proofs of those parts that have not been discussed by Cardot et al. (2007), we have used conditions (3.11)-(3.13). Conditions (3.11) and (3.12) are milder than the two conditions $\sup_j(|\langle \rho, e_j \rangle| j^{2.5}) < \infty$ and $\sup_j(\lambda_j j^4) < \infty$ used by Cardot et al. (2007).

4 Numerical results

The purpose of this simulation study is to investigate a numerical assessment of the confidence intervals for $E[Y|X = x]$. We generated random samples of sizes $n = 50$ and 200 from model (2.1), with $a = 0$ and the slope function $\rho(u) = \sum_{j=1}^{50} j^{-\beta} e_j(u)$, for $\beta = 1, 2$, or 3 , where $e_j(u) = \sqrt{2} \cos(j\pi u)$ for $j \geq 1$. The random functions X_i were distributed as $X(u) = \sum_{j=1}^{50} U_j \phi_j(u)$, where the U_j denoting independent variables with zero means and respective variances $\lambda_j = 4j^{-\alpha}$, for $\alpha = 1.5$. The distribution of the U_j was either normal $N(0, \lambda_j)$, centred standard exponential or uniform $[-\sqrt{3}, \sqrt{3}]$ with the same variance, and the errors ϵ_i were independent and normal $N(0, 1)$. We took the future observation of X to be $x(u) = \sum_{j=1}^{50} j^{-\gamma} e_j(u)$, in which $\gamma = 1, 2$, or 3 . Because in practice functional data are usually observed at discrete time points, it is good to assess the confidence intervals for prediction in this case as well as the case in which the X_i are observed continuously. In the discrete case, each X_i was observed discretely on an equally-spaced grid of 100 points with additive $N(0, 1)$ random noise. The nominal coverage level was 95%, and we computed all coverages by averaging over 5000 simulated data sets. In order to obtain the appropriate values of the cut-off point m , we use the cross-validation technique proposed by Hall and Yang (2010).

4.1 Results under different settings

Table 1 contains coverage levels and mean length of two-sided, equal-tailed asymptotic confidence intervals for both continuous and discrete cases. The table shows that in general, the obtained coverage levels increase and the mean length of confidence intervals decreases, as the sample size n increases. For example, under the centred exponential distribution with $\alpha = 1.5$, $\beta = 2$ and $\gamma = 3$, the coverage levels related to the continuous and discrete cases for $n = 50$ are 0.87 and 0.92, respectively, but they improve to 0.92 and 0.94 for $n = 200$, respectively. Table 1 also show that, the coverage levels under the normal distribution are higher than those obtained under the centred exponential and uniform distribution.

We have also carried out a simulation study with $\sigma = 2$. It can be seen that the coverage levels obtained with $\sigma = 2$ are the same as those resulted with $\sigma = 1$, but the mean length of the confidence intervals with $\sigma = 2$ is twice larger than those obtained with $\sigma = 1$.

5 Conclusions

The theoretical results in this paper are related to the proof of the asymptotic normality of the prediction in the functional linear model with functional covariate and scalar response. These results are applied to construct of the pointwise confidence intervals for the prediction in the functional linear regression model.

The simulation study also shows that the empirical coverage rates of these intervals are close to the nominal coverage even for small sample size $n = 50$. An open question for further study concerns the situations where the response is also a functional variable.

References

- Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Ann.Statist.***34**, 2159-2179.
- Cardot, H., Ferraty, F. and Sarda, P.(1999).Functional linear model. *Statist.Probab.Lett.***45**,11-22.
- Cardot, H., Mas, A. and Sarda, P. (2007). CLT in functional linear regression models. *Probab. Theory Relat. Fields* **138**, 325–361.
- Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal component analysis. *J. Roy. Statist. Soc. Ser. B* **68**, 109–126.

- Hall, P. and Hosseini-Nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. *Math. Proc. Camb. Phil. Soc.* **149**, 225-256.
- Hall, P. and Yang, Y.-J. (2010). Ordering and selecting components in multivariate or functional data linear prediction. *J. Roy. Statist. Soc. Ser. B* **72**, 93-110.
- James, G. M., Wang, J. and Zhu, J. (2009). Functional linear regression that's interpretable. *Ann. Statist.* **37**, 2083-2108.
- Li, Y. and Hsing, T. (2007). On rates of convergence in functional linear regression. *J. Multivariate Anal.* **98**, 1782-1804.
- Ramsay, J. O. and Silverman, B. (2005). *Functional data analysis*. Springer, New York.
- Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873-2903.



On the modified of generalized weighted exponential distribution

Omid Kharazmi¹

Vali-e-asr,Rafsanjan,Iran

Abstract: The new class of weighted exponential (WE) distributions obtained by Gupta and Kundu (2009) applying Azzalini's method to the exponential distribution. Kharazmi, et al (2015) extended weighted exponential distribution to the generalized weighted exponential (GWE) distribution and studied its different properties. In this study, we generalize the GWE distribution to a new class of extended generalized weighted exponential (MGWE) distribution with one scale parameter and one shape parameters. Several statistical and reliability properties of this new class of distribution are obtained. Estimation and inference procedure for distribution parameters are investigated. Finally, we show that the proposed model can provide better fit than recent class of extended weibull by using two real data examples.

Keywords Weighted exponential distribution, Hazard function, Mean residual life time, Stochastic orders, Maximum likelihood estimates.

1 Introduction

The now widely known skew-normal distribution is just one special case belonging to the family of distributions introduced by Azzalini (1985). For the first time Azzalini (1985) introduce a shape parameter to a normal distribution. Afterwards extensive work on introducing shape parameters for other symmetric distributions have been defined and several properties and their inference procedures have been discussed by Several authors, see for example, Genton (2004), Arnold and Beaver (2000a), and Nadarajah (2009). However, not much attention has been paid to introduce an extra shape parameter to a skewed distribution based on Azzalini's idea. Recently some authors effort to implement

¹Omid Kharazmi : Omid.Kharazmi@yahoo.com

Azzalini's idea for skewed distributions, the new class of weighted exponential (WE) distribution obtained by Gupta and Kundu (2009) by implementing Azzalini's method to the exponential distribution. Shakhathreh (2012) generalized the WE distribution to the two-parameter weighted exponential distributions (TWE). Kharazmi, et al (2015) extended weighted exponential distribution to the generalized weighted exponential (GWE) distribution and studied its different properties. Several interesting properties of this distribution have been established by authors [1]. GWE distribution generalizes the WE and TWE distributions and contains these distributions as its sub-models. The aim of this paper is to introduce a class of modified generalized weighted exponential (MGWE) distribution. A proposed MGWE distribution provides the GWE, WE and TWE distributions as its sub-models. This extension offers more flexible distributions with applications in reliability engineering and life time modeling. Also, several properties of the MGWE distribution have been established.

2 Definition and basic properties

The weighted exponential distribution was introduced in the seminal paper by Gupta and Kundu (2009). A random variable is said to have weighted exponential distribution, denoted by $WE(\alpha, \lambda)$, if its probability density function (PDF) is given as

$$f_x(x, \alpha, \lambda, n) = \frac{\alpha + 1}{\alpha} \lambda e^{-\lambda x} (1 - e^{-\lambda \alpha x})^n. \quad (2.1)$$

where $x > 0, \alpha > 0, \beta > 0$. Here α and λ are the shape and scale parameters, respectively. The main properties and different interpretations of this density are established by authors. Recently, Kharazmi et al. generalized WE distribution to the generalized weighted exponential distribution $GWE(\alpha, \lambda, n)$ with the following PDF

$$f_X(x, \alpha, \lambda, n) = \frac{\alpha}{B(\frac{1}{\alpha}, n + 1)} \lambda e^{-\lambda x} (1 - e^{-\lambda \alpha x})^n. \quad (2.2)$$

The corresponding cumulative distribution function (CDF) for $x \geq 0$, becomes

$$F_X(x, \alpha, \lambda, n) = 1 - \frac{\alpha}{B(\frac{1}{\alpha}, n + 1)} \sum_{j=0}^n \frac{(-1)^j \binom{n}{j}}{\alpha j + 1} e^{-\lambda(\alpha j + 1)x}. \quad (2.3)$$

Also, survival function , failure rate function and moment generating function have been established respectively as

$$\bar{F}_X(x, \alpha, \lambda, n) = \frac{\alpha}{B(\frac{1}{\alpha}, n+1)} \sum_{j=0}^n \frac{(-1)^j \binom{n}{j}}{\alpha j + 1} e^{-\lambda(\alpha j + 1)x}. \quad (2.4)$$

$$h(x, \alpha, \lambda, n) = \frac{\lambda e^{-\lambda x} (1 - e^{-\lambda \alpha x})^n}{\sum_{j=0}^n \frac{(-1)^j \binom{n}{j}}{\alpha j + 1} e^{-\lambda(\alpha j + 1)x}}. \quad (2.5)$$

$$M_X(t) = \sum_{j=0}^n \frac{(-1)^j \binom{n}{j}}{\lambda(\alpha j + 1) - t}. \quad (2.6)$$

where the beta function is defined in the usual way as $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ and $x > 0, \alpha > 0, \beta > 0, n \geq 1$.

Definition 2.1. A random variable X is said to have modified generalized weighted exponential distribution, denoted by $EGWE(\alpha, \lambda, n, m)$, if its probability density function (PDF) is given as. with integer $n \geq 1, m \geq 1$, shape parameters α, β and scale parameter λ , if the PDF of X is given as following

$$f_X(x, \alpha, \lambda, n, m) = \frac{\alpha}{[B(\frac{1}{\alpha}, n+1) - B(\frac{1}{\alpha}, n+m+1)]} \lambda e^{-\lambda x} (1 - e^{-\lambda \alpha x})^n [1 - (1 - e^{-\lambda \alpha x})^m]. \quad (2.7)$$

where the beta function is defined in the usual way as $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ and $x > 0, \alpha > 0, \beta > 0, n \geq 1$.

3 Construction of extended generalized weighted exponential (EGWE) distribution

To construct GWE distribution we use Balakrishnan's idea, about generalized skew-normal distribution, see Balakrishnan(2002), Arnold and Beaver (2002), by replacing skew-normal with weighted exponential density. Let $X, X_1, \dots, X_n, Y_1, \dots, Y_m$ be a random sample of size $n+m+1$ from the exponential distribution with scale parameter λ then we have

$$X | \text{Max}(X_1, X_2, \dots, X_n) \leq \alpha X \leq \text{Max}(Y_1, Y_2, \dots, Y_m) \sim MGWE(\alpha, \lambda, n, m). \quad (3.1)$$

This method of construction is known as a selection model. Arellano et al. (2006). Also, there are different ways to construct the EGWE distribution especially, the ones which are provided for the WE distribution by Gupta and Kundu (2009). The following methods provide some of these known ways in generalizing distribution families. The proofs of all methods are straightforward except method 3.4, it is explored as a main theorem in section 5.

Method 3.1.

The concept of weighted distributions is important in a wide range of statistical applications. A density function g is said to be a weighted density function corresponding to density function f with weight $w > 0$

$$g(x) = \frac{w(x)}{E[w(X)]} f(x), \quad x \geq 0 \quad (3.2)$$

where $0 < E[w(X)] < \infty$, Patil and Rao (1977, 1978). The MGWE distribution can be obtained as a special form of the weighted distribution by taking weighted function

$$w(x, \alpha) = [1 - e^{-\lambda\alpha x}]^n [1 - (1 - e^{-\lambda\alpha x})^m] \quad (3.3)$$

Method 3.2.

The MGWE distribution can be obtained as mixture of two GWE distribution. Mixture of life time distributions have been used extensively in reliability and survival analysis, and they have also been generalized by allowing negative mixing weights, which arise naturally under the formation of some structures of reliability systems. These models provide flexible distributions for modeling dependent lifetimes from heterogeneous populations. let $w = \frac{B(\frac{1}{\alpha}, n+1)}{[B(\frac{1}{\alpha}, n+1) - B(\frac{1}{\alpha}, n+m+1)]}$ Then MGWE distribution can be obtained as

$$f(x) = wf_U(x) + (1 - w)f_V(x) \quad (3.4)$$

where $U \sim GWE(\alpha, \lambda, n)$ and $U \sim GWE(\alpha, \lambda, n + m)$, $\alpha > 0, \lambda > 0$.

4 Statistical and reliability properties (MGWE) distribution

In this section we study the some statistical and reliability properties of the MGWE distribution, such as the distribution function (CDF), survival function (SF), failure rate (or hazard) function (FR),

moment generating function (MGF) and kth moment.

4.1 Distribution function, survival, and failure rate function

By using the method 3.2 CDF of (3) can be written as

$$F(x) = wF_U(x) + (1 - w)F_V(x) \quad (4.1)$$

where $U \sim GWE(\alpha, \lambda, n)$, $U \sim GWE(\alpha, \lambda, n + m)$ and $w = \frac{B(\frac{1}{\alpha}, n+1)}{[B(\frac{1}{\alpha}, n+1) - B(\frac{1}{\alpha}, n+m+1)]}$.

Also, survival function (SF) is given by

$$\bar{F}(x) = w\bar{F}_U(x) + (1 - w)\bar{F}_V(x) \quad (4.2)$$

From (3.4) and (4.2) it is easy to verify that the failure rate function is given by

$$h(x) = wh_U(x) + (1 - w)h_V(x) \quad (4.3)$$

where $U \sim GWE(\alpha, \lambda, n)$, $U \sim GWE(\alpha, \lambda, n + m)$ and $w = \frac{B(\frac{1}{\alpha}, n+1)}{[B(\frac{1}{\alpha}, n+1) - B(\frac{1}{\alpha}, n+m+1)]}$. The failure rate is a key notion in reliability and survival analysis for measuring the ageing process. Understanding the shape of the failure rate is important in reliability theory, risk analysis and other disciplines. The concepts of increasing and decreasing failure rates for univariate distributions have been found very useful in reliability theory. The classes of distributions having these ageing properties are designated as the IFR and DFR distributions, respectively. The independent random variables U and V are IFR see Kharazmi et al (2015). So the mixing these variables can produce divergent class of life time distribution. Fig. 1 shows some failure rate function shapes for some values of α , $\lambda = 1$, $n = 10$, $m = 1$. The failure rate function increases.

4.2 Moment generating function

Now let us consider different moments of the $MGWE(\alpha, \lambda, n, m)$ distribution. Some of the most important features and characteristics of a distribution can be studied through its moments, such as moment generating function. The moment generating function of form (3.4) is immediately written as

$$M_X(t) = wM_U(t) + (1 - w)M_V(t) \quad (4.4)$$

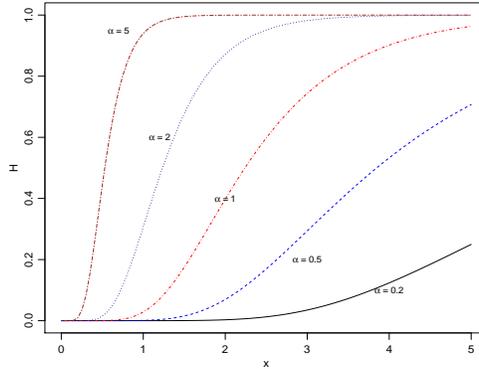


Figure 1: Hazard function of the MGWE distribution for different values of α .

where $U \sim GWE(\alpha, \lambda, n)$, $U \sim GWE(\alpha, \lambda, n + m)$ and $w = \frac{B(\frac{1}{\alpha}, n+1)}{[B(\frac{1}{\alpha}, n+1) - B(\frac{1}{\alpha}, n+m+1)]}$.

5 Main results

In this section we proposed main results about the MGWE family distribution.

Theorem 5.1. 1.If $n \rightarrow \infty$, then MGWE random variable is degenerated at point 0.

2. If $\alpha \rightarrow 0$, then $f_X(x, \alpha, \lambda, n, m)$ converges to $\text{Gamma}(n + 1, \lambda)$

3. If $\alpha \rightarrow \infty$, then $f_X(x, \alpha, \lambda, n, m)$ converges to $\text{Exp}(\lambda)$

4. Let $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$ be a sample from exponential distribution with mean $\frac{1}{\lambda}$

let $U = \text{Min}(X_1, X_2, \dots, X_n)$, $V = \text{Min}(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)$ then by mixing random variables U and V with weight $w = \frac{B(\frac{1}{\alpha}, n)}{[B(\frac{1}{\alpha}, n+1) - B(\frac{1}{\alpha}, n+m)]}$ resulted in $MGWE(\alpha = 1, \lambda, n - 1, m - 1)$

Proof. By consider Kharazmi et,al (2015) results, the proofs of all cases are straightforward. □

6 Data analysis

Strength Data

Strength data measured in GPA, the single carbon fibers, and impregnated 1,000-carbon fiber tows. Single fibers were tested under tension at gauge length 1 mm. The data are provided below:

2.247 2.64 2.908 3.099 3.126 3.245 3.328 3.355 3.383 3.572 3.581 3.681 3.726 3.727 3.728 3.783 3.785
 3.786 3.896 3.912 3.964 4.05 4.063 4.082 4.111 4.118 4.141 4.246 4.251 4.262 4.326 4.402 4.457 4.466
 4.519 4.542 4.555 4.614 4.632 4.634 4.636 4.678 4.698 4.738 4.832 4.924 5.043 5.099 5.134 5.359 5.473
 5.571 5.684 5.721 5.998 6.06

See Kundu et al.(2014). We have used the proposed MGWE model and estimates of α , λ are 0.004, and 5.811 respectively. The corresponding log-likelihood value and Aikake’s Information Criterion (AIC) are -68.38 and 140.76 respectively . The KS distance between the fitted and the empirical distribution functions is 0.058 and the associated pvalue is 0.9996. Table 1 shows some comparison with other distributions.

Table 1:

Distribution	Log-likelihood	AIC	K-S distance	pvalue
Gamma	-71.8824	147.7648	0.0973	0.6410
Weibull	-70.3395	144.679	0.0648	0.9726
GE	-74.6607	153.3214	0.1221	0.3735
MOGE	-67.8507	141.7014	0.0474	0.9996

7 Conclusions

In this paper, we have proposed the new class of modified weighed exponential distribution denoted by MGWE. The proposed distribution generalizes the WE and GWE distributions and contains these distributions as its sub-models. One application of the MGWE distribution to real data set is provided to illustrate that this distribution provides a better fit than its sub-models.

References

Azzalini, A. (1985).A class of distributions which includes the normal ones, Scandinavian Journal of Statistics**12**, 171–178.

Gupta, R.D. and Kundu, D.(2009). A new class of weighted exponential distributions, *Statistics* **43**, 621–634.

Kharazmi, O., Mahdaiv, A. and Fathizadeh, M.(2015). Generalized weighted exponential distribution, *Communication in Statistics-Simulation and Computation* **44**, 1557–1569, 2015.

Shakhatreh, M. K.(2012). A two-parameter of weighted exponential distributions, *Statistics and Probability Letters* **82**, 252-261,2012.



A New Bayesian Unit Root Test in Unobserved-ARCH Models

Fazlolah Lak¹, Mahmood Afshari, Behzad Gholizadeh
Department of Statistics, Faculty of Sciences
Persian Gulf University, Boushehr 75168, Iran

Abstract: In this paper a new posterior odds analysis is proposed to test for a unit root in Unobserved-ARCH models. Our analysis extends the Bayesian unit root test of . Also a method for estimating the probability of the null hypothesis in prior odds ratio is demonstrated. Simulation study shows that this method is efficient. An empirical study, based on time series data of daily exchange rate of the German Marc with respect to the Greek Drachma, is applied using this method.

Keywords: Bayes factor, Gibbs sampling, Monte Carlo Markov Chain, Unobserved-ARCH model.

Mathematics Subject Classification (2010): 62M10, 62F15.

1 Introduction

Financial markets sometimes appear quite calm and at other times highly volatile. The key of describing how this volatility changes over time, is the testing of the null hypothesis that the observations available on a variable or their deviations from some appropriate constants follow a random walk. Explicit models of heteroskedasticity have a long history in statistics and econometrics. Engle (1982) propose Autoregressive Conditional heteroskedasticity (ARCH) class of models spurred a virtual arms race into the development of new and better procedures for modeling and forecasting time-varying financial market volatility. Nelson (1991), Bollerslev (2008) characterized this general class of models to ensure the success of it in various applications in economics and finance models. One of the members of the class of ARCH models is Unobserved Autoregressive Conditional Heteroscedasticity (U-ARCH). In U-ARCH model, the ARCH component is observed with error, or, it may be considered as a latent process. Redaction of the unit root null hypothesis in U-ARCH results that observations

¹Speaker Name: Fazlolah Lak Lak@pgu.ac.ir

have a statistical model. The literature that considers U-ARCH models is extensive; see [Giakoumatos et al \(2005\)](#), [Lak \(2011\)](#) for detailed surveys.

The purpose of this paper is to provide a new bayesian method for testing the unit root null hypothesis in U-ARCH time series by expanding the method in [Lak \(2011\)](#). Simulation studies show that the new method is more powerful than the method provided by [Lak \(2011\)](#).

The structure of this paper is as follows. In section 2, the U-ARCH model is introduced. Some properties of U-ARCH model and sampling method by using the Metropolis-Hastings algorithm is illustrated. The computation of posterior odds ratio by using Monte Carlo method is explained in section 3. A new prior distribution, will be offered in section 4, there a method for estimating the prior probability of the null hypothesis along with the results of the simulation studies is proposed. Finally in section 5, the new methodology is used with real data.

2 The Unobserved ARCH Model and Posterior Odds Ratio

In Unobserved ARCH model, the ARCH components are observed with errors. By using hierarchical structure of conditional densities, one can write this model by following form:

$$y_t|f_t, \sigma^2 \sim N(f_t, \sigma^2), \quad f_t|f_{t-1}, \alpha, \beta, f_0 \sim N(0, h_t), \quad h_t = \alpha + \beta f_{t-1}^2. \quad (2.1)$$

where, y_1, \dots, y_T is a realization of the process. f_t is Unobserved ARCH component at time t and f_0 is initial state or the "history" of the unobserved components, and $N(\cdot, \cdot)$ is the normal distribution. Because, h_t is the variance of a normal distribution, it must be positive. For this purpose, the parameters α and β are restricted to be positive. The additional restriction $0 < \beta \leq 1$ is imposed so that the ARCH component of the model is covariance stationary.

For applying MCMC methods, transformations $g = \sqrt{\frac{\alpha}{\beta}}$ and $w_t = \sqrt{\frac{\beta}{\alpha}} f_t$ are used for $t = 0, 1, \dots, T$. Therefore, the likelihood density is equal to:

$$\begin{aligned} L(g, \beta, \sigma^2|y_1, \dots, y_T, w_0, \dots, w_T) &\propto \frac{g^2}{\sigma^T \beta^{T/2-1} \prod_{t=1}^T \sqrt{1 + w_{t-1}^2}} \\ &\times \exp \left\{ \frac{1}{2\beta} \sum_{t=1}^T \frac{w_t^2}{1 + w_{t-1}^2} \right\} \end{aligned}$$

$$- \frac{1}{2} \left[\frac{1}{\sigma^2} \sum_{t=1}^T (y_t - gw_t)^2 + \frac{(gw_0)^2}{\nu} \right] \Bigg\},$$

where $g, \sigma > 0$ and $0 < \beta < 1$.

Due to the fact that parameter β is restricted to lie within $(0, 1)$, we suggest two prior distribution for β : the uniform distribution $\pi(\beta) = 1$ on $(0, 1)$ which is a non-informative prior distribution and truncated inverse gamma within $(0, 1)$, as a natural conjugate prior distribution, formulated as $\pi(\beta) \propto \frac{1}{\beta^{a_1+1}} \exp\{-\frac{b_1}{\beta}\} I_{0 < \beta < 1}$. The improper prior distribution $\pi(g) = \frac{1}{g^2}$ and the conjugate inverse gamma $IG(a_2, b_2)$ are considered for g and σ^2 respectively. By using inverse gamma prior distribution for β , the posterior distribution of parameters takes the form:

$$\begin{aligned} \pi(g, \beta, \sigma^2, w_0, \dots, w_T | y_1, \dots, y_T) &\propto \frac{1}{(\sigma^2)^{\frac{T}{2}+a_2} \beta^{\frac{T}{2}+a_1} \prod_{t=1}^T \sqrt{1+w_{t-1}^2}} \\ &\times \exp \left\{ -\frac{1}{2\beta} \sum_{t=1}^T \frac{w_t^2}{1+w_{t-1}^2} - \frac{b_1}{\beta} \right. \\ &\left. - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - gw_t)^2 - \frac{b_2}{\sigma^2} - \frac{gw_0^2}{\nu} \right\}. \end{aligned} \quad (2.2)$$

The posterior distribution does not have a regular form. Therefore, one cannot directly collect a sample from it and estimate the parameters. In this paper, the Gibbs sampling method suggested by [Giakoumatos et al \(2005\)](#) is used for the sampling of the joint posterior distribution of all the unknown parameters. Given the joint posterior distribution, the full conditional densities for inverse gamma prior distribution of β are:

- $\sigma^2 | \cdot \sim IG \left(\frac{T+2a_2-2}{2}, \frac{1}{2} \sum_{t=1}^T (y_t - gw_t)^2 + b_2 \right),$
- $\beta | \cdot \sim IG \left(\frac{T+2a_1-2}{2}, \frac{1}{2} \sum_{t=1}^T \frac{w_t^2}{1+w_{t-1}^2} + b_1 \right),$
- $g | \cdot \sim N \left(\frac{\nu \sum_{t=1}^T w_t y_t}{m}, \frac{\sigma^2 \nu}{m} \right) I(g \geq 0), \quad m = \sigma^2 w_0^2 + \nu \sum_{t=1}^T w_t^2,$
- $w_0 | \cdot \propto N \left(0, \frac{\nu}{g^2} \right) \frac{1}{\sqrt{1+w_0^2}} \exp \left\{ -\frac{1}{2\beta} \frac{w_0^2}{1+w_0^2} \right\},$
- $w_t | \cdot \propto N \left(m_t, s_t^2 \right) \frac{1}{\sqrt{1+w_t^2}} \exp \left\{ -\frac{1}{2\beta} \frac{w_{t+1}^2}{1+w_t^2} \right\},$

- $w_T|\cdot \sim N(m_T, s_T)$,

where $m_t = \frac{y_t g \beta (1+w_{t-1}^2)}{g^2 \beta (1+w_{t-1}^2) + \sigma^2}$, and $s_t^2 = \frac{\sigma^2 \beta (1+w_{t-1}^2)}{g^2 \beta (1+w_{t-1}^2) + \sigma^2}$. Also for uniform prior density of β , take $a_1 = b_1 = 0$. The full conditional density of w_t for $t = 0, 1, \dots, T-1$ does not have known forms. One way to deal with it is to use Metropolis-Hastings steps which allow us to sample non-standard densities. For detail study see [Robert and Casella \(2004\)](#).

Our aim is to perform hypothesis testing $\beta < 1$ against $\beta = 1$, which can be thought as comparing the two competing models. Let M_0 be the model formulated in the null hypothesis, M_1 the model formulated under the alternative hypothesis, $\pi(M_k)$ the prior model probability, $p(y|M_k)$ and $p(M_k|y)$ the marginal likelihood and the posterior probability density respectively, where $k = 0, 1$. By assuming the prior probabilities ϱ_0 and ϱ_1 for H_0 and H_1 hypotheses respectively, the associated Bayes estimator, is:

$$\phi^\pi(y) = \begin{cases} 1, & \text{if } P^\pi(\beta \in \Theta_0|y) > \varrho_1, \\ 0, & \text{otherwise.} \end{cases}$$

where $\Theta_0 = \{1\}$ and $\Theta_1 = \{\beta|\beta < 1\}$. In other words, H_0 will be rejected if:

$$\frac{P(M_0|y)}{P(M_1|y)} = \frac{P(y|M_0)}{P(y|M_1)} \times \frac{\varrho_0}{\varrho_1} > 1.$$

For this class of losses, the null hypothesis H_0 is rejected when the posterior probability of H_0 is too small. Notice that ϕ^π only depends on ϱ_0/ϱ_1 and that the larger ϱ_0/ϱ_1 is, i.e., the more important a wrong answer under H_0 is relative to H_1 , the smaller the posterior probability of H_0 needs to be for H_0 to be accepted.

When the flat prior is used i.e. $\varrho_0 = \varrho_1 = 0.5$, the posterior odds ratio will be the equal bayes factor (BF) that is defined as the ratio of marginal likelihoods under the respective hypotheses. As mentioned above, M_0 is favored over M_1 if the posterior odds ratio is larger than 1. In this paper, the logarithm of the bayes factor will be employed. If the sign of \log_{10}^{BF} is positive, model M_0 is favored over M_1 , [Robert \(2007\)](#). To calculate the marginal likelihood, the method presented by [Chib \(1995\)](#) will be applied, which is based on the identity in θ ;

$$m_k(y) = \frac{f_k(y|\theta)\varrho_k}{\pi_k(\theta|y)}, \quad k = 0, 1,$$

where $\pi_k(\theta|y)$ is the posterior density of θ under H_k . As advised by Chib (1995), we set θ at its posterior mean of the Gibbs sample and estimate the likelihood function of the Unobserved-ARCH model. The posterior value of $\pi_k(\theta|y)$ is computed utilizing the method suggested by Chib (1995).

3 A New Prior Distribution

In this section, two models are compared with posterior odds ratio computed using the MCMC methods. In fact, the bayes factor is only defined when $\varrho_0 \neq 0$ and $\varrho_1 \neq 0$. This implies that, if H_0 or H_1 are a prior impossible, the observations will not modify this absolute information. Therefore, the point null-hypothesis $H_0 : \beta = 1$ cannot be tested under a continuous prior distribution. So the testing of the point-null hypothesis imposes a drastic modification of the prior distribution. Indeed, to test the hypothesis $\beta = 1$ implies that there is a chance that β truly is equal to 1 and therefore that some possibly defined indication of this fact has been provided. Thus consideration of the probability ρ for the model M_0 means that a prior distribution has a positive mass at unity. So the prior distribution of β appears as the following mixture distribution:

$$\pi(\beta) = \rho I(\beta = 1) + (1 - \rho) f_U(\beta) I(0 < \beta < 1). \quad (3.1)$$

In this distribution, assume ρ is a bernoulli variable with the probability $p = p(\rho = 1)$. Also assume U is a continue random variable on $(0, 1)$, and U is independent of ρ . This type of priors is widely used in the literature of unit root test and tends to favor stationary or trend-stationary hypotheses, Robert (2007).

As mentioned earlier, when flat prior probabilities on Θ_0 and Θ_1 is used, the posterior odds ratio takes the same value as the Bayes factor. Therefore, only the Bayes factor is used for comparison of bayesian models. Hence, the power of the unit root test can not be improved. Thus, to increase the power of the test, we try to set a prior distribution for models.

Regard to the new prior distribution for β , the full conditional distribution of β is equal to:

$$\beta|\cdot = \begin{cases} 1, & w.p. \ p, \\ IG(\frac{T}{2} + a_1 - 1, \frac{1}{2} \sum \frac{w_t^2}{1+w_t^2} + b_1), & w.p. \ 1 - p, \end{cases}$$

where $\beta|\cdot$ is the posterior density for β , while, Inverse Gamma distribution is used as f_U and for uniform prior, $a_1 = b_1 = 0$. Since the full conditional density of β is a mixture distribution, β , with the probability ρ , equals to 1, and, with the probability $1 - \rho$, has the posterior distribution of $IG(\frac{T}{2} - 1 + a_1, \frac{1}{2} \sum \frac{w_i^2}{1+w_i^2-1} + b_1)$.

Given the influence that sample size exerts on the accuracy of the method proposed by Lak (2011), in order to determine p , first β will be estimated using Lak’s method and then ρ , using the value of $1 - \hat{\beta}$ and sample size, will be determined. In this paper, the recommended values for p are 0.10, $\frac{100}{n} \exp\{-\hat{\beta}\}$ and $\frac{100}{2n} \exp\{-\hat{\beta}\}$, where $\hat{\beta}$ is an estimate of β by MCMC method. In studying the effect of value of ρ on accuracy of estimation of β , a series of simulations experiments are carried out to address these three methods. Moreover, three values of β are considered, namely 1, 0.98, and 0.95. The choice of σ^2 and α are motivated by the results in the empirical section of this paper and most of the estimates of σ^2 and α have values around 0.01 and 0.005 respectively. The time series of length $n = 500$, $n = 1000$, and $n = 1500$ are simulated for each of them. The Gibbs sampling takes out 10000 times the first 2000 generated samples of which are burn-in. All experiments are duplicated 100 times. Table 1 gives the average value of the posterior mean of β and the logarithm of POR in

Table 1: Comparisons of the mean proportional odds ratio when $p = p(\rho = 1)$.

Candidate values		$n = 500$			$n = 1000$		
β	p	$\hat{\beta}$	$\tilde{\beta}$	\log_{10}^{POR}	$\hat{\beta}$	$\tilde{\beta}$	\log_{10}^{POR}
0.95	0.10	0.9401	0.9713	-0.9873	0.9491	0.9849	-2.9815
	$\frac{100}{n} \exp\{-\hat{\beta}\}$		0.9608	-1.3210		0.9615	-3.0159
	$\frac{100}{2n} \exp\{-\hat{\beta}\}$		0.9443	-1.7269		0.9519	-3.8971
0.98	0.10	0.9739	0.9852	-1.2546	0.9789	0.9861	-1.7421
	$\frac{100}{n} \exp\{-\hat{\beta}\}$		0.9821	-1.7140		0.9837	-1.9327
	$\frac{100}{2n} \exp\{-\hat{\beta}\}$		0.9783	-1.9851		0.9817	-2.1158
1	0.10	0.9879	0.9923	2.0018	0.9892	0.9928	1.7148
	$\frac{100}{n} \exp\{-\hat{\beta}\}$		0.9919	1.4139		0.9922	1.4972
	$\frac{100}{2n} \exp\{-\hat{\beta}\}$		0.9911	1.6511		0.9916	1.7690

100 replications. The results of simulation study, when the value of $p(\rho = 1) = p$ is $\frac{100}{2n} \exp\{-\hat{\beta}\}$, are more accurate and the values of POR in stationary time series cases are larger. Moreover, in the

Table 1:Continued Table 1.

Candidate values		$n = 1500$		
β	p	$\hat{\beta}$	$\tilde{\beta}$	\log_{10}^{POR}
0.95	0.10	0.9503	0.9852	-3.5471
	$\frac{100}{n} \exp\{-\hat{\beta}\}$		0.9604	-4.2856
	$\frac{100}{2n} \exp\{-\hat{\beta}\}$		0.9602	-6.1873
0.98	0.10	0.9793	0.9869	-3.5487
	$\frac{100}{n} \exp\{-\hat{\beta}\}$		0.9830	-2.3934
	$\frac{100}{2n} \exp\{-\hat{\beta}\}$		0.9810	-2.8102
1	0.10	0.9896	0.9941	1.4897
	$\frac{100}{n} \exp\{-\hat{\beta}\}$		0.9924	1.5678
	$\frac{100}{2n} \exp\{-\hat{\beta}\}$		0.9921	1.7721

proposed method, the null hypothesis is truly accepted 99 times, which substantiates the power of new method compared to Lak's method in which the null hypothesis is truly accepted only 93 times. Also in the majority of the cases, where the null hypothesis is truly accepted, the logarithm of proportional odds ratio came to be higher. It should be noticed that, $\tilde{\beta}$ is fakery estimated upper than real value while p is equal to 0.10, which shows that using p , (dependent to sample size) get better results in the estimation of β .

4 Empirical Study

In the empirical study, the daily exchange rate of the Germany Marc with respect to the Greek Drachma is studied. Data are daily exchange rates recorded from December 16th 1993 to May 2nd 1997. To implement the Gibbs sampling, three dispersed initial values for all the parameters are utilized. The Gibbs Sampling was run (in R software) to obtain 30,000 samples. The first 5,000 burn-in iterations were discarded, and the last 25,000 iterates were used to estimate all unknown parameters. This process was repeated 100 times. Based on the results from simulation studies, we assume $p = \frac{100}{2n} \exp\{-\hat{\beta}\}$. Also, the same set of flat prior distributions was used for α and σ^2 . For the parameter β , the $U(0, 1)$, and $IG(2.0017, 0.95)$ priors were used.

The posterior mean of unknown parameters are given in Table 2. Monte Carlo estimates posterior

odds ratios are smaller than 1, due to fact that the posterior odds favour stationarity.

Table 2: *The posterior mean and odds ratios of unknown parameters.*

Prior	σ^2	α	β	\log_{10}^{POR}
Uniform	0.089	0.0041	0.9781	-1.9936
Truncated Inverse gamma	0.091	0.0046	0.9803	-3.8067

5 Conclusion

A new prior distribution for unit root testing of econometric time series was introduced. It was shown that, the power of new unit root test and the accuracy of this method in estimating parameters, is better than the traditional method that proposed by Lak (2011). Also, a method for estimating p in the new prior of unit root test was proposed. The proposed method was applied for testing and estimating parameters with real data.

References

- Bollerslev, T. (2008). Glossary to arch (garch). *Creates Research Paper*, 49.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90(13)**, 1332.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimation of the variance of the United Kingdom inflation. *Econometrica* **50**,9871007.
- Giakoumatos, S. G., Delaportas, P., Politis, D. N. (2005). Bayesian analysis of the unobserved ARCH model. *Statistics and Computing*, 103-111.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**, 347-370.
- Lak, F. (2011). Bayesian Unit Root Testing in Unobserved-ARCH Models, *Communications in Statistics-Simulation and Computation*, **40(2)**, 208-215.

Li, Y. and Yu, J. (2010). A New Bayesian Unit Root Test in Stochastic Volatility Models. *Research Collection School of Economics, Singapore Management University*. 1240, 2010.

Robert, C.P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.

Robert, C.P. and Casella, G. (2004). *Monte Carlo statistical methods*, Springer Science & Business Media.



α -power transformation method for generating distributions

Abbas Mahdavi⁵

Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.

Abstract: A new way of introducing a parameter to expand a family of distributions is introduced and as an application a new two-parameter extension of the exponential distribution is proposed and studied. Various fundamental properties including explicit expressions for the moments, quantiles, mode and moment generating function of the new distribution are derived. The maximum likelihood estimator of unknown parameters cannot be obtained in explicit forms and must be obtained by solving a one-dimensional optimization procedure. One real data analysis has been presented for illustrative purposes.

Keywords: exponential distribution, hazard rate function, α -power transformation method, maximum-likelihood estimation, survival reliability function.

Mathematics Subject Classification (2010): 62E10 60E05.

1 Introduction

Adding parameters to an existing distribution to expand family of distributions is a very common approach for developing more flexible models. Azzalini (1985) introduced the method of generating skew distributions to adding a skew parameter to normal distribution. This method combines two independent symmetric distributions to form a skewed distribution. Let X and Y be continuous independent random variable symmetric about 0, with PDFs f and g and CDFs F and G . Then for any $\lambda \in \mathfrak{R}$, the function

$$2f(x)g(\lambda x), \quad x \in \mathfrak{R}. \quad (1.1)$$

is a valid PDF that developed by the method of generating skew distributions. Mudholkar and Srivastava (1993) proposed the exponentiated method to introduce an additional parameter to Weibull

⁵Abbas Mahdavi : a.mahdavi@vru.ac.ir

distribution. In this method an additional parameter plays the role of a power for the cumulative distribution function (CDF) of an existing distribution. Suppose $F(x)$ be the CDF of a continuous random variable X , the CDF of the exponentiated random variable X defined as

$$F_E(x) = F^\alpha(x), \quad x \in \mathfrak{R}, \alpha > 0. \quad (1.2)$$

Eugene, Lee et al. (2002) proposed the beta generated method that uses the beta distribution with shape parameters α and β as the generator to develop the beta generated distributions. The CDF of a beta-generated random variable X is defined as

$$G(x) = \int_0^{F(x)} b(t) dt, \quad (1.3)$$

where $b(t)$ is the PDF of the beta random variable and $F(x)$ is the CDF of any random variable X . For more survey about methods to generating distributions see Lee, Famoye et al. (2013). The aim of this paper is to introduce a method to generate a family of distributions based on an existing CDF, called α -power transformation (APT). As an application we generated a two-parameter (scale and shape) α -power exponential family (APE). It is observed that APE distribution has several desirable properties and it can be used as an alternative to the several two-parameter life time models. The rest of the paper is organized as follows. In Section 2, we introduce APT method to generate distributions. In Section 3, we define the APE distribution. A range of mathematical properties of the new distribution is considered in Sections 4-5. These include quantile function, simulation, mode, moments. The MLEs and their properties are discussed in Section 6. In Section 7, we provide one real data analysis result. Finally, concluding remarks are provided in Section 8.

2 α -power transformation method

Let $F(x)$ be the CDF of continuous random variable X . Then then the *alpha*-power transformation of $F(x)$ for $x \in R$, is defined as follows:

$$F_{APT}(x) = \frac{\alpha^{F(x)} - 1}{\alpha - 1}, \quad x \in \mathfrak{R}, \alpha > 0, \alpha \neq 1. \quad (2.1)$$

The probability density function (PDF) corresponding to the APT family in (2.1) is defined as

$$f_{APT}(x) = \frac{\log \alpha}{\alpha - 1} f(x) \alpha^{F(x)}, \quad (2.2)$$

where $f(x)$ is the PDF of random variable X . It can be shown that when $\alpha \rightarrow 1$ then $f_{APT}(x)$ converge to $f(x)$.

3 α -power exponential distribution

In this section, first we provide the definition of α -power exponential distribution by taking $F(x)$ as a CDF of exponential distribution with mean $1/\lambda$ in (2.1) and then we study different properties of the APE distribution.

Definition 3.1. *The random variable X is said to have APE distribution denoted by $APE(\alpha, \lambda)$, with the shape and scale parameters as $\alpha > 0, \alpha \neq 1$ and $\lambda > 0$, respectively, if the PDF of X is*

$$f(x) = \frac{\log \alpha}{\alpha - 1} \lambda e^{-\lambda x} \alpha^{1-e^{-\lambda x}}, \quad x > 0. \quad (3.1)$$

The CDF of APE distribution is given by

$$F(x) = \frac{\alpha^{(1-e^{-\lambda x})} - 1}{\alpha - 1}, \quad x > 0. \quad (3.2)$$

Also, the survival reliability function $S(x)$ and the hazard rate function (HRF) $h(x)$ for APE distribution are in the following forms

$$S(x) = \frac{\alpha}{\alpha - 1} (1 - \alpha^{-e^{-\lambda x}}), \quad x > 0, \quad (3.3)$$

$$h(x) = \frac{\lambda e^{-\lambda x} \alpha^{-e^{-\lambda x}} \log \alpha}{1 - \alpha^{-e^{-\lambda x}}}, \quad x > 0. \quad (3.4)$$

Figure 1 shows some of the different shapes of $APE(\alpha, \lambda)$ and HRF for selected values of α and fixed scale parameter $\lambda = 1$. It is an unimodal density function and for fixed scale parameter as the shape parameter increases it is becoming more and more symmetric. If $\alpha > 1$, the hazard function is a non-decreasing function and converge to λ , and if $\alpha < 1$ it is a non-increasing function and converge to λ . Therefore the hazard function of the APE distribution behaves like the hazard function of the gamma distribution with the following PDF

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad \alpha, \lambda, x > 0. \quad (3.5)$$

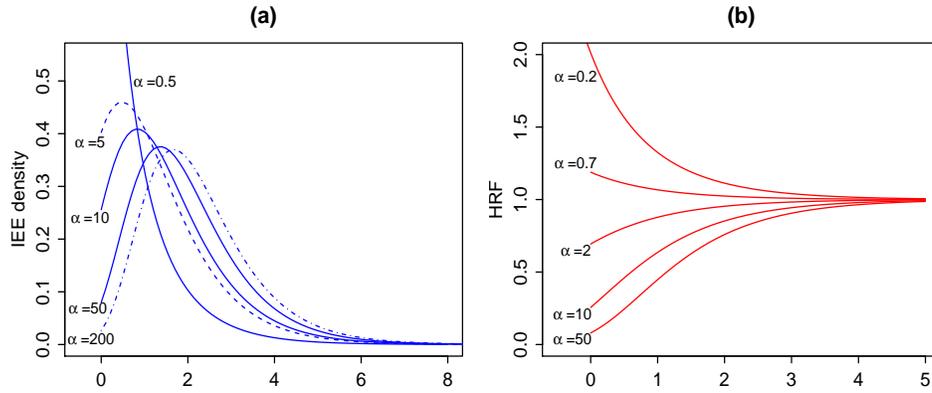


Figure 1: (a) and (b) the PDF and HRF of APE distribution with various shape parameter and fixed scale parameter $\lambda = 1$.

(a)

APE distribution has an explicit expression of the distribution function or the survival function. Therefore, it can be used as an alternative to a gamma distribution and it might work better than a gamma distribution in some situations.

4 Quantile function and simulation

The quantile function of APE distribution is given by

$$x_p = \frac{-1}{\lambda} \log\left\{\frac{\log \frac{\alpha}{(\alpha-1)^{p+1}}}{\log \alpha}\right\}. \tag{4.1}$$

The median can be derive from (4.1) by letting $p = 0.5$ as

$$m = \frac{-1}{\lambda} \log\left\{\frac{\log \frac{2\alpha}{\alpha+1}}{\log \alpha}\right\}. \tag{4.2}$$

Simulation of APE random variable follows directly from Equation (4.1) by

$$X = \frac{-1}{\lambda} \log\left\{\frac{\log \frac{\alpha}{(\alpha-1)U+1}}{\log \alpha}\right\}, \tag{4.3}$$

where U is a standard uniform variate.

5 Mode and moments

The mode of APE exist for any $\alpha > e$ in point of

$$M = \frac{\log(\log \alpha)}{\lambda}. \quad (5.1)$$

Using the series representations

$$\alpha^{-u} = \sum_{k=0}^{\infty} \frac{(-\log \alpha)^k u^k}{k!}, \quad (5.2)$$

and

$$\int u^n (\log u)^m du = u^{n+1} \sum_{k=0}^m (-1)^k \frac{m! (\log u)^{m-k}}{(m-k)! (n+1)^{k+1}}, \quad (5.3)$$

we derive two closed-form expressions for the moment-generating function (MGF) and for the n th moment as

$$M_X(t) = \frac{\lambda \alpha}{1 - \alpha} \sum_{k=0}^{\infty} \frac{(-\log \alpha)^{k+1}}{(k\lambda - t + \lambda)k!}, \quad t < \lambda, \quad (5.4)$$

$$E(X^n) = \frac{\alpha n!}{\lambda^n (1 - \alpha)} \sum_{k=1}^{\infty} \frac{(-\log \alpha)^k}{k^n k!}. \quad (5.5)$$

6 Maximum likelihood estimation

In this section, we consider the estimation of the unknown parameters by the method of maximum likelihood. Let x_1, x_2, \dots, x_n be a sample from $APE(\alpha, \lambda)$ distribution. Then, the loglikelihood function based on the given random sample is

$$\begin{aligned} \log L = & n \log \alpha - n \log(\alpha - 1) + n \log \lambda - \lambda \sum_{i=1}^n x_i \\ & - (\log \alpha) \sum_{i=1}^n e^{-\lambda x_i} + n \log(\log \alpha). \end{aligned} \quad (6.1)$$

The first order derivatives of $\log L$ are

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} - \frac{n}{\alpha - 1} - \frac{1}{\alpha} \sum_{i=1}^n e^{-\lambda x_i} + \frac{n}{\alpha \log \alpha}, \quad (6.2)$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i + (\log \alpha) \sum_{i=1}^n x_i e^{-\lambda x_i}. \quad (6.3)$$

The system of equations obtained by setting these derivatives to zero does not have a closed solution and must be solved numerically. In this paper we use the optim function from the statistical software R (R Development Core Team, 2013) to maximize the logarithm of the likelihood function. The solution of this maximization problem are the maximum likelihood estimators (MLEs) of the parameters, and are denoted respectively by $\hat{\alpha}$ and $\hat{\lambda}$. In this section, we use an uncensored data set corresponding to intervals in days between 109 successive coal-mining disasters in Great Britain, for the period 1875-1951, published by Maguire, Pearson et al. (1952). The sorted data are given as follows: 1 4 4 7 11 13 15 15 17 18 19 19 20 20 22 23 28 29 31 32 36 37 47 48 49 50 54 54 55 59 59 61 61 66 72 72 75 78 78 81 93 96 99 108 113 114 120 120 120 123 124 129 131 137 145 151 156 171 176 182 188 189 195 203 208 215 217 217 217 224 228 233 255 271 275 275 275 286 291 312 312 312 315 326 326 329 330 336 338 345 348 354 361 364 369 378 390 457 467 498 517 566 644 745 871 1312 1357 1613 1630. We compare APE distribution with four other two-parameter models

- gamma distribution with PDF (3.5)
- Weibull distribution with PDF $f(x) = \alpha \lambda (\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha}$, $\alpha, \lambda, x > 0$.
- exponentiated exponential distribution (EE) introduced by Gupta, Gupta et al. (1998) with PDF $f(x) = \alpha \lambda (1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}$, $\alpha, \lambda, x > 0$.
- weighted exponential distribution (WE) proposed by Gupta and Kundu (2009) with PDF $f(x) = \frac{\alpha+1}{\alpha} \lambda e^{-\lambda x} (1 - e^{-\alpha \lambda x})$, $\alpha, \lambda, x > 0$.

To see which one of these models is more appropriate to fit coal-mining data the MLEs of parameters, Kolmogorov-Smirnov statistics and p-values are given in Table 1. Comparing the log-likelihood values and p-values based on the Kolmogorov Smirnov test, we see that APE distribution provides a significantly better fit than the other four models. The relative histogram and the fitted APE distribution are plotted in Figure 2 (a). In order to assess if the model is appropriate, the plots of the fitted APE survival function and empirical survival function are displayed in Figure 2 (b).

Table 1: The maximum likelihood estimates and Kolmogorov-Smirnov statistics and p-values for coal-mining data.

The model	MLEs of the parameters	Log-likelihood	K-S statistic	p-value
gamma	$\hat{\alpha} = 0.8555, \hat{\lambda} = 0.0037$	-702.4007	0.0823	0.4517
Weibull	$\hat{\alpha} = 0.8848, \hat{\lambda} = 0.0046$	-701.7724	0.0784	0.5135
EE	$\hat{\alpha} = 0.8605, \hat{\lambda} = 0.0039$	-702.5524	0.0830	0.4402
WE	$\hat{\alpha} = 35.2748, \hat{\lambda} = 0.0045$	-705.1641	0.0836	0.4313
APE	$\hat{\alpha} = 0.2807, \hat{\lambda} = 0.0030$	-701.2132	0.0742	0.5852

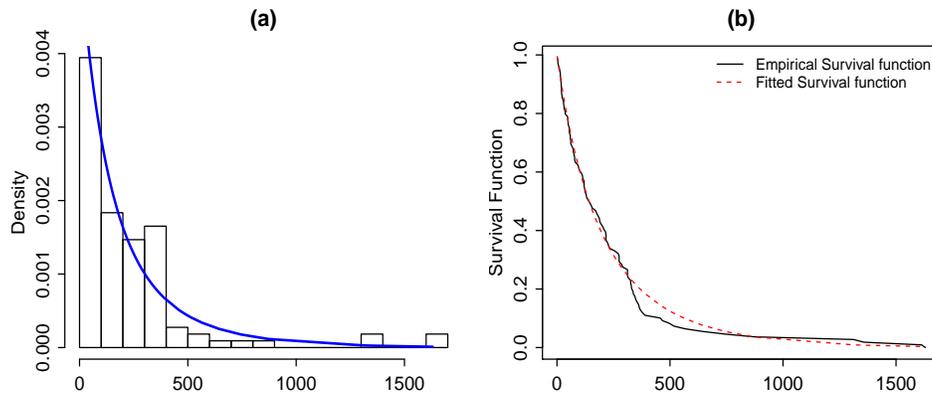


Figure 2: (a) the histogram and the fitted APE distribution. (b) the fitted APE survival function and empirical survival function for coal-mining data.

Conclusion

A new method to generate families of distributions is introduced. As an application a new two-parameter family of distributions was provided by using the particular case that F is the CDF of exponential distribution which may sometimes be a competitor to the Weibull, gamma and other two-parameter life time models. The flexibility of the proposed distribution and increased range of skewness was able to fit and capture features in one real data set much better than some popular distributions. Also the generation of random samples from the APE is quite straight forward, that is useful to perform the simulation experiments or parametric bootstrapping. Finally, future works is needed to implement and investigate the behavior of the proposed method to expand other family of

distributions. It should be mentioned that although we have used the APT method to the exponential distribution function, similar method can be used to the normal distribution also. It will introduce a skewness parameter to a normal distribution similar to the skew-normal distribution of Azzalini (1985). It will be interesting to study the properties of this skew-normal distribution and also to develop the inferential procedures. The work is in progress, it will be reported later.

References

- Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian journal of statistics*, **12**, 171-178.
- Eugene, N., et al. (2002). Beta-normal distribution and its applications, *Communications in statistics-Theory and methods*, **31(4)**, 497-512.
- Gupta, R. C., et al. (1998). Modeling failure time data by Lehman alternatives, *Communications in statistics-Theory and methods*, **27(4)**, 887-904.
- Gupta, R. D. and D. Kundu. (2009). A new class of weighted exponential distributions, *Statistics*, **43(6)**, 621-634.
- Lee, C., et al. (2013). Methods for generating families of univariate continuous distributions in the recent decades, *Wiley Interdisciplinary Reviews: Computational Statistics*, **5(3)**, 219-238.
- Maguire, B., et al. (1952). The Time Intervals Between Industrial Accidents, *Biometrika*, **39(1-2)**, 168-180.
- Mudholkar, G. S. and D. K. Srivastava. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data, *IEEE Transactions on Reliability*, **42(2)**, 299-302.
- Team, R. C. (2005). R: A language and environment for statistical computing, *R foundation for Statistical Computing*.



A goodness-of-fit test for the extreme value distribution in the presence of outliers

Abbas Mahdavi, Hossein Negarestani⁶

Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran

Abstract: This article provides a simple robust method to test of goodness of fit for the extreme value distribution (Type I family) by using the new diagnostic tool called the "Forward Search" (FS) method. One of the powerful goodness-of-fit tests is the correlation coefficient test, but this test suffers from the presence of outliers. We introduce the FS version of this test that is not affected by the outliers. Also by use the transformation study, an application to the two-parameter Weibull distribution is investigated. The performance and the ability of this procedure to capture the structure of data are illustrated by some simulation studies.

Keywords Forward search procedure, Goodness of fit test, Correlation coefficient test, Outlier, Robust approach.

Mathematics Subject Classification (2010): 62G10 62J20.

1 Introduction

The extreme value distributions are widely used in risk management, finance, insurance, economics, hydrology, material sciences, telecommunications, and many other industries dealing with extreme events. One of the basic tools that is useful to testing goodness of fit is the correlation coefficient goodness-of-fit test based on quantile-quantile (QQ) plot. The purpose of this article is to adopt the Forward Search (FS) method in the goodness of fit test for extreme value distribution. One of the powerful tests is the correlation coefficient test introduced by Filliben (1975) for testing normality. Kinnison (1989) assessed the goodness of fit of Type-I extreme value distribution based on Filliben's correlation coefficient test and examined its power properties for various alternative models. The

⁶Hossein Negarestani : negarestani@vru.ac.ir

correlation coefficient statistic is not a robust statistic and hence the presence of outliers influences this test strongly. The test involves computing the correlation coefficient between the ranked data and the expected value of the order statistic with the same rank. In this paper we try to determine how many and which observations agree with the hypothesis of extreme value distribution and also as an application for testing the goodness of fit for Weibull distribution. The FS approach is a powerful general method that provides diagnostic plots for finding outliers and discovering their underlying effects on models fitted to the data and for assessing the adequacy of the model. Atkinson and Riani (2000, 2002, 2004) developed the FS procedure for regression modeling and multivariate analysis frameworks. The FS method starts from a small, robustly chosen subset of the data. The method increases the subset size by using some measure of closeness from fitted model until finally all the data are fitted. The outliers enter the model in the last steps and the entrance point of the outliers can be revealed by monitoring some statistics of interest during the process. The paper is organized as follows. In Section 2 we briefly introduce correlation coefficient goodness-of-fit test for the extreme value distribution. Section 3 presents the proposed forward search algorithm in correlation coefficient goodness-of-fit test for the extreme value distribution. In Section 4, we describe the easy and fast method for estimating the null distribution of proposed test statistic. In Section 5 the performance of the method is illustrated by simulating some datasets. Concluding remarks are provided in Section 6.

2 Correlation coefficient goodness-of-fit test for the extreme value distribution

The correlation coefficient test was introduced by Filliben (1975) to testing goodness of fit for the normal distribution. Kinnison (1989) assessed the goodness of fit of Type-I extreme value distribution based on Filliben's correlation coefficient test and examined its power properties for various alternative models. A QQ plot is a common and basic technique used for finding a suitable model to data. When comparing an observed data to a hypothesized distribution, the plot of the ordered observations versus the appropriate quantiles of assumed distribution, should look approximately linear and hence the product moment correlation coefficient which measures the degree of linear association between two random variables is an appropriate test statistic.

The correlation coefficient goodness-of-fit test for the extreme value distribution is built as follows. Let X be a random variable from the extreme value distribution

$$G(x) = 1 - \exp(-\exp[(x - \mu)/\sigma]), \quad (2.1)$$

where μ and σ are unknown location and scale parameters. In such location-scale model there is a simple relationship between the p -quantiles of X and $W = (X - \mu)/\sigma$ the standard extreme value variable ($\mu = 0, \sigma = 1$). The p -quantile of X , defined by $P(X \leq x_p) = p$, is

$$x_p = \mu + \sigma \log(-\log(1 - p)). \quad (2.2)$$

Thus x_p is a linear function of $w_p = \log(-\log(1 - p))$, the p -quantile of W .

Let $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ be an ordered sample of size n from X , for appropriate $p_i; i = 1, 2, \dots, n$ (the plotting positions), we can approximate x_{p_i} by the i -th order sample $X_{(i)}$. Thus the correlation coefficient statistic, R , for goodness-of-fit test is defined as the correlation between ordered sample $X_{(i)}$ and the p_i -quantile of W , w_{p_i} . Many plotting positions have been proposed in the literature, in this paper we use the median rank due to its robustness property and is therefore used in the case of skewed distributions, such as the extreme value distribution (see D'Agostino et al., 1986 and Castillo et al., 2005). The median rank, $m_{(i)}$, of the i -th order statistics is given by

$$m_{(i)} = B_{0.5, i, n-i+1}, \quad (2.3)$$

where $B_{0.5, \alpha, \beta}$ is the median of the beta distribution with parameters α and β .

The distribution of R can be estimate by means of Monte Carlo simulation for different sample sizes. One rejects the hypothesized distribution (1) if the observed value of R is smaller than the critical value.

As an application, to use the correlation coefficient test for testing the validity of two-parameter Weibull distribution to the data with cumulative distribution function (cdf):

$$f(x; \alpha, \beta) = 1 - \exp[-(x/\beta)^\alpha]. \quad (2.4)$$

It is necessary to transform into an extreme value distribution with location parameter $\mu = \log(\beta)$ and scale parameter $\sigma = 1/\alpha$ by taking the log of the data.

3 Forward search in correlation coefficient goodness-of-fit test for the extreme value distribution

Let $\mathbf{x}_{(\cdot)} = (x_{(1)}, x_{(2)}, \dots, x_{(n)})$ the vector of ordered observations comes from an extreme value distribution (1), then it is possible to write

$$x_{(i)} = \mu + \sigma w_{m_i} + \varepsilon_i, \quad (3.1)$$

where $w_{m_i} = \log(-\log(1 - m_i))$ and m_i is the median rank defined in (3). In this section we use the FS method introduced by Atkinson and Riani (2000) to analysis the behavior of regression model (5). The FS method is a powerful approach not only to detecting outliers, but also for investigating their effect on the estimation of parameters and on aspects of inference about models. The basic idea of the FS approach is to order the observations by their closeness to the fitted model. The FS method is made up of the following three main steps: the starting point is to find the appropriate starting subset of observations, the second step presents the plan to progressing in FS and the last step is monitoring some suitable quantities during the search. In the following subsections we describe how these three points are performed.

3.1 Step 1: Choice of the initial subset

Starting point of the FS procedure is choosing outlier free subset of observations robustly. To start the FS approach, the size of initial subset must be specified. That size can be as small as $p = 2$. Therefore, we search over subsets of p observations to find the best subset of observations. The initial subset can be achieved by use of robust regression estimator Least Median of Squares regression estimator (LMS) proposed by Rousseeuw (1984).

3.2 Step 2: Progressing in the search

At each step of the search, the procedure adds to the subset the observation that is closest to the previously fitted model. Let $S^{(k)}$ be a subset of size k , the FS moves to $S^{(k+1)}$ in the following way: after the least square regression model is fitted to the $S^{(k)}$ subset, all observations are ordered according

to their square residuals, now we choose the $k + 1$ observations with the smallest square residuals. This procedure is repeated until all observations are entered into the model.

3.3 Step 3: Monitoring the search

For detecting and determining the effect of outliers, some statistics of interest must be monitor during the search. The FS version of correlation coefficient, \mathbf{R}_{FS} , is defined as a collection of R (correlation coefficient) statistics computed for different subsets of $\mathbf{x}_{(\cdot)}$ and corresponding units of w_{m_i} . Let $w^{(k)}$ be the units of w_{m_i} corresponding to the subset $S(k)$, then the R_{FS} is defined as

$$R_{FS} = (R_{S^{(p)},w^{(p)}}, \dots, R_{S^{(k)},w^{(k)}}, \dots, R_{S^{(n)},w^{(n)}}). \quad (3.2)$$

The empirical quantiles of (6) during the search can be estimated by simulation in each step of the search. In any step of the search the acceptance region lies between the value of the chosen quantile an 1.

4 Simulating the distribution

The empirical null distribution of (6) can be found by simulating numerous samples generated from a standard extreme value distribution. Since the FS is a reiterative algorithm, this way of estimating distribution is very time consuming. Atkinson and Riani (2006) proposed the method of "Ordered Observations" to estimate the distribution of outlier test statistic. In the following subsection we describe this simple and fast method briefly.

The FS orders all observations in each steps of the search. In the absence of outliers, when we move from $S^{(k)}$ to $S^{(k+1)}$, most of the time just one new observation joins the subset and this ordering does not change much during the search. Hence, we can order the observations just one according to square residuals resulting from the chosen initial subset, denoted by $\mathbf{x}_{(ord)}$. In the step k of the search we just choose the first k observations of $\mathbf{x}_{(ord)}$.

Figure 1 shows the 5% bounds of the empirical distribution and the estimated distribution using the ordered observations method for sample sizes $n = 50$ and $n = 100$.

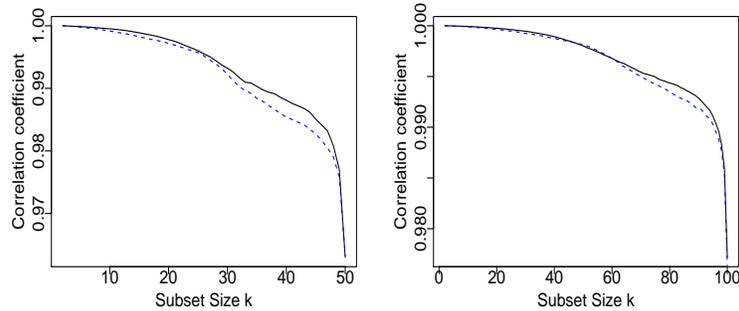


Figure 1: The 5% bounds of the empirical distribution (continues lines) and the estimated distribution using the ordered observations method (dashed lines) for sample sizes (left panel) and (right panel).

5 Simulation study

In order to evaluate the proposed procedure we conduct simulation studies that aim to consider the behavior of statistic (6) in the presence of outliers and ability of FS for detecting them. Consider six samples are generated in the following way:

- Sample A: 100 observations are generated from a standard extreme value distribution.
- Sample B: 95 observations are generated from a standard extreme value distribution and for contamination 5 observations are generated from a $N(\mu = 3, \sigma = 1)$.
- Sample C: 95 observations are generated from a standard extreme value distribution and for contamination 5 observations are generated from a $Uniform(a = 3, b = 4)$.
- Sample D: 100 observations are generated from a $Weibull(sh = 2, sc = 2)$.
- Sample E: 95 observations are generated from a $Weibull(sh = 2, sc = 2)$ and for contamination 5 observations are generated from a $N(\mu = 10, \sigma = 1)$.
- Sample F: 95 observations are generated from a $Weibull(sh = 2, sc = 2)$ and for contamination 5 observations are generated from a $Uniform(a = 9, b = 10)$.

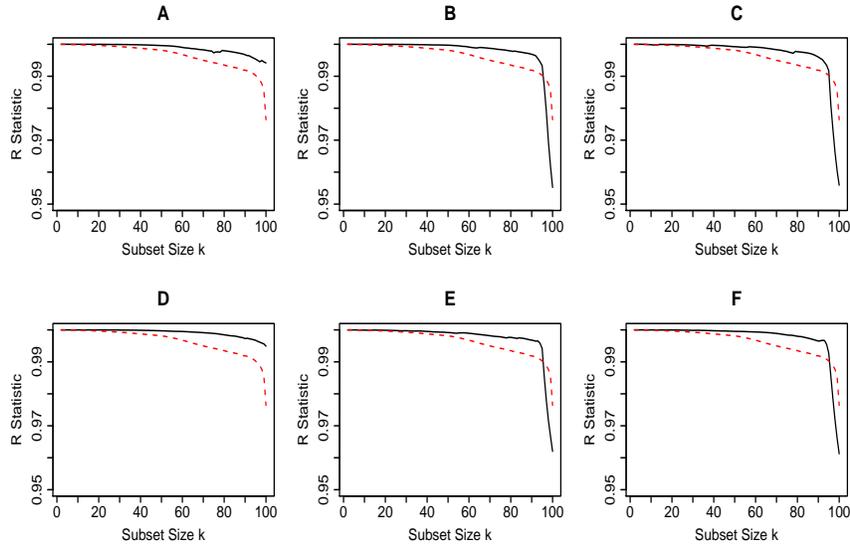


Figure 2: Forward plots of R_{FS} during the search for samples A-F.

In Figure 2, the values of R_{FS} during the search are plotted for samples A-F and compared with estimated corresponding 5(dashed line) of its distribution obtained from ordered observations described in Section 4. The null hypothesis of extreme value distribution is accepted in each step of the search for clean sample A and it is rejected after entrance of outliers in the last steps (step 96 onwards) for contaminated samples B and C indicating 5 observations are outliers. Also, the same results for samples D-F summarized as: the null hypothesis of two-parameter Weibull distribution is accepted in any step of the search for clean sample D and it is rejected from step 96 onwards for samples E and F.

6 Concluding remarks

In this paper, a new robust method to testing the goodness of fit for extreme value distribution has been presented. The approach gives information about the distribution of majority of the data and the percentage of contamination. At every step of the FS, the proposed statistic is computed and with a graphical approach a cut-off point divides the group of outliers from the other observations. In order to illustrate the application and the advantage of the FS approach we used some artificial examples.

We have used the simple and fast method to find distribution of proposed statistic. Furthermore, we showed an application of the proposed approach to the goodness-of-fit test for the two-parameter Weibull distribution.

References

- Atkinson, A. C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, Springer, New York.
- Atkinson, A. C. and Riani, M. (2002), Forward search added-variable t-tests and the effect of masked outliers on model selection, *Biometrika*, **89**(4), 939-946.
- Atkinson, A. C. and Riani, M. (2004), The Forward search and data visualisation, *Computational Statistics*, **19**, 29-54.
- Atkinson, A. C. and Riani, M. (2006), Distribution theory and simulations for tests of outliers in regression, *Journal of Computational and Graphical Statistics*, **15**, 460-476.
- Atkinson, A. C., Riani, M. and Cerioli, A. (2010), The forward search: theory and data analysis, *Journal of the Korean Statistical Society*, **39**, 117-134.
- Bertaccini, B. and Varriale, R. (2007), Robust analysis of variance: An approach based on the forward search, *Computational statistics and data analysis*, **51**, 5172-5183.
- Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005), *Extreme value and related models with applications in engineering and science*, John Wiley and Sons, Hoboken.
- Coin, D. (2008), Testing normality in the presence of outliers, *Statistical Methods and Applications*, **17**, 3-12.
- D'Agostino, R. B. and Stephens, M. A. eds. (1986), *Goodness of fit techniques*, Marcel Dekker, New York.
- Filliben, J. J. (1975), The probability plot correlation coefficient test for normality, *Technometrics*, **17**, 111-117.

Kinnison, R. (1989), Correlation coefficient goodness of fit test for extreme value distribution, *The American Statistician*, **43**, 98-100.

Rousseeuw P.J. (1984), Least median of squares regression, *Journal of the American Statistical Association*, **79**, 871-880.



A note on Bayesian inference for the Topp-Leone distribution

M. Mahdizadeh⁷

Department of Statistics, Hakim Sabzevari University, P.O. Box 397, Sabzevar, Iran

Abstract: This article deals with parameter estimation for the Topp-Leone distribution using Bayesian approach. Especial attention is given to its performance relative to traditional maximum likelihood method. This does not seem to be attended enough in the literature. Monte Carlo simulations are employed to compare different estimators based on complete and censored samples.

Keywords: Asymmetric loss function; Failure rate; Unit interval.

Mathematics Subject Classification (2010): 62N01; 62F15.

1 Introduction

The Topp-Leone distribution is a univariate continuous distribution which was first proposed and applied as a model for some failure data by [Topp and Leone \(1955\)](#). A random variable X is said to have Topp-Leone distribution with parameter θ , denoted by $X \sim TL(\theta)$, if its pdf can be expressed as

$$f(x) = \theta(2 - 2x)(2x - x^2)^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0, \quad (1.1)$$

where θ is the shape parameter. The corresponding cdf is also given by

$$F(x) = (2x - x^2)^\theta, \quad 0 < x < 1. \quad (1.2)$$

The Topp-Leone distribution with the range $(0, 1)$ can be used as an alternative distribution to beta distribution which has the same range.

The Topp-Leone distribution has received much attention in recent years. [Nadarajah and Kotz \(2003\)](#) provided a motivation for this distribution based on its hazard rate function and then studied the moments and some other main properties of the Topp-Leone distribution. [Ghitany et al. \(2005\)](#)

⁷ mahdizadeh.m@live.com

studied some reliability measures and their stochastic orderings for the Topp-Leone model. [Genc \(2012\)](#) discussed the moments of order statistics coming from this distribution. [Sindhu et al. \(2013\)](#) focused on Bayesian estimation of θ based on type II censored data. [Feroze and Aslam \(2013\)](#) considered Bayesian estimation of θ based on complete and (singly and doubly) type II censored data. In the last two works, performance of the maximum likelihood estimator (MLE) has not been compared with Bayesian estimators. We plan to address this topic in this study.

Section 2 reviews MLEs and Bayesian estimators of θ using complete and singly type II censored data. Results of Monte Carlo experiments conducted to assess performance of the proposed estimators are reported in Section 3. Lastly, we conclude in Section 4.

2 The estimators

2.1 Complete data

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a sample of size n from $TL(\theta)$. The associated likelihood function is of the form

$$L(\theta|\mathbf{x}) \propto \theta^n e^{-T\theta}, \quad (2.1)$$

where $T = \sum_{i=1}^n \log(2x_i - x_i^2)^{-1}$. Differentiating (2.1) with respect to θ , and equating the result to zero, the MLE is obtained as $\hat{\theta}_{ML} = n/T$.

To proceed with Bayesian methodology, one needs to specify a prior distribution reflecting the available knowledge about the parameter. To this end, we consider a simple family of improper distributions

$$g(\theta) \propto \theta^{-2c}, \quad c > 0; \theta > 0, \quad (2.2)$$

which includes some well-known priors. In particular, $c = 0$ leads to the uniform prior, and $c = 0.5$ gives Jeffreys prior for $TL(\theta)$. The latter is defined as $g(\theta) \propto \sqrt{I(\theta)}$, where $I(\theta)$ is the Fisher information about θ contained in the sample.

Using (2.1) and (2.2), the posterior distribution turns out to be

$$g(\theta|\mathbf{x}) = \frac{T^{n-2c+1}}{\Gamma(n-2c+1)} \theta^{n-2c} e^{-T\theta}, \quad \theta > 0, \quad (2.3)$$

where $\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} du$. The above density function can be used to derive Bayes estimators under different loss functions. In the sequel, we consider three choices.

The squared error loss (SEL) function is defined as $L_1(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ which is symmetric in the sense that it assigns equivalent dimensions to underestimation and overestimation. Under SEL function, Bayes estimator and the corresponding risk are given by

$$\hat{\theta}_{SE} = \frac{1}{T}(n - 2c + 1), \quad r(\hat{\theta}_{SE}) = \frac{1}{T^2}(n - 2c + 1).$$

The use of symmetric loss functions, however, may not be appropriate in many real situations, especially when underestimation and overestimation have different consequences. In such situations, the use of asymmetric loss functions are suggested. The precautionary loss (PL) function, introduced by [Nostrom \(1996\)](#), enjoys this property and is defined as $L_2(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2 / \hat{\theta}$. Under PL function, Bayes estimator and the corresponding risk are given by

$$\hat{\theta}_{PL} = \frac{1}{T} \sqrt{(n - 2c + 1)(n - 2c + 2)}, \quad r(\hat{\theta}_{PL}) = \frac{2}{T} \left(\sqrt{(n - 2c + 1)(n - 2c + 2)} - (n - 2c + 1) \right).$$

Another well-known alternative is the linear-exponential loss (LL) function defined (in a special case) as $L_3(\hat{\theta}, \theta) = \exp\{(\hat{\theta} - \theta)\} - (\hat{\theta} - \theta) - 1$. See [Varian \(1975\)](#) for more details. Under LL function, Bayes estimator and the corresponding risk are given by

$$\hat{\theta}_{LL} = -(n - 2c + 1) \log\left(\frac{T}{T + 1}\right), \quad r(\hat{\theta}_{LL}) = (n - 2c + 1) \left(\frac{1}{T} + \log\left(\frac{T}{T + 1}\right) \right).$$

2.2 Censored data

Let we observe n items for possible failure and only first $m (< n)$ failure times have been observed. That is, $x_1 < \dots < x_m$, and the remaining $n - m$ items are still working. Under the assumptions that the lifetimes of the items are iid from $TL(\theta)$, the associated likelihood function is of the form

$$L(\theta|\mathbf{x}) \propto \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \theta^m e^{-\phi_k \theta}, \tag{2.4}$$

where $\phi_k = \sum_{i=1}^m \log(2x_i - x_i^2)^{-1} + k \log(2x_m - x_m^2)^{-1}$. Differentiating (2.4) with respect to θ , and equating the result to zero, we arrive at

$$\frac{m}{\theta} + \sum_{i=1}^m \log(2x_i - x_i^2)^{-1} - (n - m) \log(2x_m - x_m^2) \frac{(2x_m - x_m^2)^\theta}{1 - (2x_m - x_m^2)^\theta} = 0 \tag{2.5}$$

Clearly, (2.5) does not admit explicit solution, and hence MLE has to be found numerically.

To derive Bayes estimators, we employ the same family of priors, and loss functions mentioned in the previous subsection. Using (2.4) and (2.2) it follows that the posterior distribution is

$$g(\theta|\mathbf{x}) = \frac{1}{D} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \theta^{m-2c} e^{-\phi_k \theta}, \quad (2.6)$$

where $D = \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \Gamma(m-2c+1) / \phi_k^{m-2c+1}$.

Under SEL function, Bayes estimator and the corresponding risk are given by

$$\hat{\theta}_{SE} = \frac{1}{D} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \frac{\Gamma(m-2c+2)}{\phi_k^{m-2c+2}}$$

and

$$r(\hat{\theta}_{SE}) = \frac{1}{D} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \frac{\Gamma(m-2c+3)}{\phi_k^{m-2c+3}} - \left[\frac{1}{D} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \frac{\Gamma(m-2c+2)}{\phi_k^{m-2c+2}} \right]^2.$$

Under PL function, Bayes estimator and the corresponding risk are given by

$$\hat{\theta}_{PL} = \left[\frac{1}{D} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \frac{\Gamma(m-2c+3)}{\phi_k^{m-2c+3}} \right]^{1/2}$$

and

$$r(\hat{\theta}_{PL}) = 2 \left\{ \left[\frac{1}{D} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \frac{\Gamma(m-2c+3)}{\phi_k^{m-2c+3}} \right]^{1/2} - \frac{1}{D} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \frac{\Gamma(m-2c+2)}{\phi_k^{m-2c+2}} \right\}.$$

Finally, Bayes estimator and the corresponding risk, under LL function, are given by

$$\hat{\theta}_{LL} = -\log \left\{ \frac{\sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} (\phi_k + 1)^{-(m-2c+1)}}{\sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \phi_k^{-(m-2c+1)}} \right\}$$

and

$$r(\hat{\theta}_{LL}) = \frac{1}{D} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \frac{\Gamma(m-2c+2)}{\phi_k^{m-2c+2}} + \log \left\{ \frac{\sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} (\phi_k + 1)^{-(m-2c+1)}}{\sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k} \phi_k^{-(m-2c+1)}} \right\}.$$

3 Simulation results

A simulation study was conducted to get insight of the finite sample behavior of the suggested estimators. For each combination of sample sizes $n = 10, 20$ and $\theta = 0.75, 1.5$, a total number of 10,000 samples were generated from $TL(\theta)$. The estimators were computed from each sample and their mean squared errors (MSEs) were determined. The bayes estimators were derived under the three loss functions mentioned in Section 2, and using prior $g(\theta) \propto \theta^{-2c}$, for $c = 0, 0.5, 0.75, 1, 2$. In estimating MSEs, the Bayes estimators are treated as classical estimators. Also, risk of Bayes estimators were estimated from 10,000 samples. The results for complete (censored) data are given in Table 1 (2). In Table 2, it is assumed that 30% of samples are censored, i.e. $m = 0.7n$. The entries corresponding to $\theta = 0.75$ ($\theta = 1.5$) are placed in the upper (lower) panel of each table. Also, efficiency of any bayes estimator relative to MLE, i.e. ratio of the corresponding MSEs, appears in parentheses.

The simulation results, not reported here, show that the shape parameter is overestimated using SEL and PL functions, while it is underestimated under LL function. The values of posterior risk is inversely proportional to sample size, and is directly proportional to true parameter value.

It is to be noted that the Bayes estimator under SEL function and Jeffreys prior is the same MLE in the case of complete data. For priors other than uniform, the Bayesian estimators are generally more efficient than the corresponding MLEs. This is more pronounced at small sample size. Also, the overall performance of LL function seems satisfactory.

Conclusion

This article attends to Bayesian methods for parameter estimation under the TL model. A general family of prior distributions, including uniform and Jeffreys priors as special cases, is considered. The Bayes estimators under SEL, PL and LL functions are derived. The estimators are compared with MLE, in terms of MSE, for some configurations of the prior distribution parameter, sample size and the shape parameter. The results confirm the preference of Bayesian inference in small sample size.

Table 1: Estimated Bayes risks and REs of different estimators for complete data and $\theta = 0.75, 1.5$

n	Loss	c				
		0	0.5	0.75	1	2
10	SE	0.0860 (0.704)	0.0788 (1.000)	0.0737 (1.170)	0.0706 (1.342)	0.0549 (1.349)
	P	0.0816 (0.593)	0.0816 (0.845)	0.0811 (1.004)	0.0814 (1.179)	0.0806 (1.463)
	L	0.0402 (0.916)	0.0368 (1.284)	0.0344 (1.457)	0.0330 (1.608)	0.0256 (1.331)
20	SE	0.0347 (0.815)	0.0330 (1.000)	0.0321 (1.089)	0.0312 (1.153)	0.0280 (1.168)
	P	0.0391 (0.729)	0.0390 (0.907)	0.0390 (1.001)	0.0389 (1.085)	0.0389 (1.212)
	L	0.0169 (0.931)	0.0160 (1.120)	0.0156 (1.201)	0.0152 (1.247)	0.0136 (1.162)

n	Loss	c				
		0	0.5	0.75	1	2
10	SE	0.3501 (0.704)	0.3133 (1.000)	0.2972 (1.171)	0.2808 (1.338)	0.2204 (1.358)
	P	0.1644 (0.594)	0.1630 (0.845)	0.1626 (1.004)	0.1624 (1.178)	0.1616 (1.473)
	L	0.1536 (1.167)	0.1377 (1.571)	0.1306 (1.727)	0.1235 (1.795)	0.0968 (1.271)
20	SE	0.1384 (0.815)	0.1296 (1.000)	0.1291 (1.088)	0.1240 (1.154)	0.1113 (1.142)
	P	0.0781 (0.729)	0.0774 (0.912)	0.0782 (1.001)	0.0777 (1.086)	0.0777 (1.193)
	L	0.0654 (1.053)	0.0613 (1.221)	0.0610 (1.304)	0.0586 (1.311)	0.0526 (1.099)

References

- Feroze, N., and Aslam, M. (2013). On Bayesian Analysis of Failure Rate under Topp Leone Distribution using Complete and Censored Samples. *International Journal of Mathematical, Computational, Physical and Quantum Engineering*, **7**, 236-242.
- Genc, A.I. (2012). Moments of Order Statistics of Topp-Leone Distribution. *Statistical Papers*, **53**, 117-131.
- Ghitany, M.E., Kotz, S., and Xie, M. (2005). On Some Reliability Measures and their Stochastic Orderings for the Topp-Leone Distribution, *Journal of Applied Statistics*, **32**, 715-722.
- Nadarajah, S., and Kotz, S. (2003). Moments of Some J-shaped Distributions, *Journal of Applied*

Table 2: Estimated Bayes risks and REs of different estimators for censored data and $\theta = 0.75, 1.5$

		c				
n	Loss	0	0.5	0.75	1	2
10	SE	0.0868 (0.703)	0.0794 (0.998)	0.0744 (1.170)	0.0717 (1.343)	0.0547 (1.323)
	P	0.0822 (0.591)	0.0822 (0.842)	0.0817 (1.002)	0.0821 (1.178)	0.0808 (1.443)
	L	0.0405 (0.918)	0.0370 (1.277)	0.0348 (1.455)	0.0335 (1.621)	0.0256 (1.292)
20	SE	0.0348 (0.813)	0.0332 (0.999)	0.0321 (1.085)	0.0312 (1.154)	0.0281 (1.158)
	P	0.0392 (0.726)	0.0393 (0.906)	0.0391 (1.000)	0.0391 (1.086)	0.0391 (1.207)
	L	0.0169 (0.929)	0.0161 (1.120)	0.0156 (1.196)	0.0152 (1.246)	0.0136 (1.146)

		c				
n	Loss	0	0.5	0.75	1	2
10	SE	0.3435 (0.703)	0.3190 (0.998)	0.2979 (1.169)	0.2820 (1.334)	0.2219 (1.370)
	P	0.1637 (0.591)	0.1647 (0.843)	0.1633 (1.002)	0.1632 (1.175)	0.1623 (1.482)
	L	0.1508 (1.171)	0.1398 (1.575)	0.1308 (1.723)	0.1238 (1.791)	0.0973 (1.289)
20	SE	0.1393 (0.811)	0.1319 (0.999)	0.1292 (1.086)	0.1247 (1.152)	0.1121 (1.155)
	P	0.0786 (0.724)	0.0782 (0.909)	0.0784 (1.000)	0.0780 (1.085)	0.0780 (1.202)
	L	0.0658 (1.050)	0.0623 (1.230)	0.0610 (1.296)	0.0589 (1.312)	0.0530 (1.123)

Statistics, **30**, 311–317.

Norstrom, J.G. (1996), The use of precautionary loss functions in risk analysis, *IEEE Trans. Reliab.* **45**, 400–403.

Sindhu, T.N., Saleem, M., and Aslam, M. (2013). Bayesian Estimation for Topp-Leone Distribution under Trimmed Samples. *Journal of Basic and Applied Scientific Research*, **3**, 347–360.

Topp, C.W., and Leone, F.C. (1955). A Family of J-shaped Frequency Functions, *Journal of the American Statistical Association*, **50**, 209–219.

Varian, H.R. (1975). *A Bayesian Approach to Real Estate Assessment*. Amsterdam: North Holland.



A gamma-Weibull distribution: Model, properties and applications

R.S. Meshkat, H. Torabi²
Yazd University, 89175-741, Yazd, IRAN

Abstract: In this paper, a new distribution, namely, gamma-Weibull (GW) distribution is proposed and studied. Some results for its moments are provided. To estimate the model parameters, the maximum likelihood estimators and the asymptotic distribution of the estimators are discussed. Finally, in order to show the usefulness of the new distribution, application of a real data set is demonstrated.

Keywords gamma-Weibull distribution, Weibull distribution, Hazard function, Maximum likelihood estimation.

Mathematics Subject Classification (2010): 60E05, 62H10, 62H12.

1 Introduction

The Weibull distribution is a very popular distribution for modelling lifetime data and to model phenomenon with monotone failure rates. However, to model monotone hazard rates, the Weibull distribution may be an initial choice because of its negatively and positively skewed density shapes, this two-parameter model is inadequate when the true hazard shape is of unimodal or bathtub nature.

Various distributional aspects of this distribution have been investigated in literature. For instance, products and ratios of Weibull random variables were studied by [Nadarajah and Kotz \(2006\)](#). For the basic distributional properties of the Weibull distribution, techniques for the estimation of its parameters as well as numerous applications, the reader is referred to [Johnson and Kotz \(1976\)](#). For various results in connection with the generalized gamma distribution and some of its asymptotic properties, the reader can be referred to [Prentice \(1974\)](#), [Smith and Naylor \(1987\)](#) and [Meeker and Escobar \(1988\)](#).

²Speaker's Name: R.S. Meshkat r.meshkat@gmail.com

In this work a four-parameter distribution are introduced, so-called the Gamma-Weibull (GW) distribution, which because of its flexibility in accommodating all the forms of the hazard rate function seems to be an important distribution.

The rest of the paper is organized as follows:

In section 2, the new distribution is introduced. Properties of this new distribution are obtained in Section 3. In Section 4, the maximum likelihood estimation are discussed. The proposed model is applied to a data set in Section 5.

2 The gamma-Weibull distribution

The two-parameter Weibull density function is usually expressed as follows:

$$g(x; \mu, \sigma) = \frac{\mu}{\sigma} x^{\mu-1} e^{-\left(\frac{x}{\sigma}\right)^\mu}, \quad x > 0,$$

where $\sigma > 0$ is a scale parameter and $\mu > 0$ is a shape parameter and its cumulative distribution function is

$$G(x; \mu, \sigma) = 1 - e^{-\left(\frac{x}{\sigma}\right)^\mu}, \quad x > 0.$$

Considering the following gamma density function

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0.$$

We define the cdf of gamma-Weibull (GW) distribution as

$$H(x; \alpha, \beta, \mu, \sigma) = \frac{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{1 - e^{-\left(\frac{x}{\sigma}\right)^\mu}}{\beta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{1}{\beta}\right)}, \quad x > 0.$$

The corresponding pdf is given by

$$h(x; \alpha, \beta, \mu, \sigma) = \frac{\mu \left(\frac{x}{\sigma}\right)^\mu \left(\frac{1 - e^{-\left(\frac{x}{\sigma}\right)^\mu}}{\beta}\right)^\alpha \exp\left\{\frac{e^{-\left(\frac{x}{\sigma}\right)^\mu} - 1}{\beta}\right\}}{x \left(e^{\left(\frac{x}{\sigma}\right)^\mu} - 1\right) \left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{1}{\beta}\right)\right]}, \quad x > 0,$$

where $\Gamma(a, z) = \int_z^\infty t^{a-1} e^{-t} dt$ denotes an incomplete gamma function and $\beta, \sigma > 0$ are scale parameters and $\alpha, \mu > 0$ are shape parameters. A random variable X follows the GW distribution with parameters

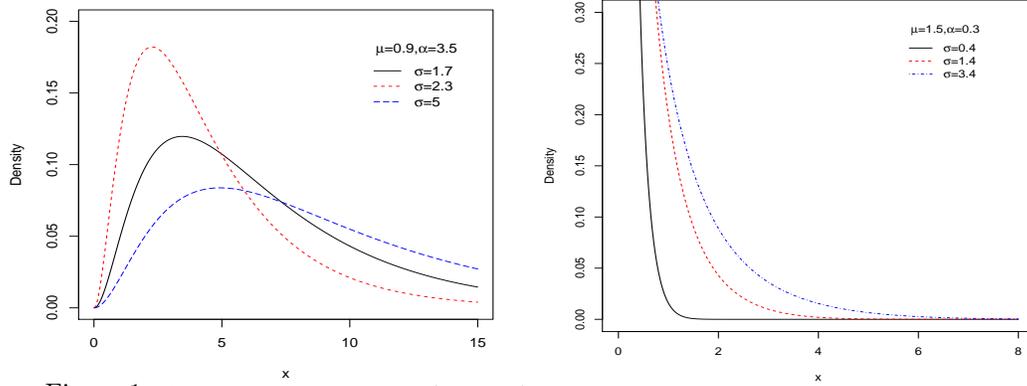


Figure 1: Plots of the pdf of the $GW(\alpha, 1, \mu, \sigma)$ distribution for some various values of parameters.

α, β, μ and σ is denoted by $X \sim GW(\alpha, \beta, \mu, \sigma)$. The associated survival and hazard hazard function of GW distribution are respectively given by

$$S(x) = \frac{\Gamma\left(\alpha, \frac{1-e^{-\left(\frac{x}{\sigma}\right)^\mu}}{\beta}\right) - \Gamma\left(\alpha, \frac{1}{\beta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{1}{\beta}\right)},$$

and

$$r(x) = -\frac{\mu\beta^{-\alpha}\sigma^{-\mu}x^{\mu-1}\left(1 - e^{-\left(\frac{x}{\sigma}\right)^\mu}\right)^\alpha \exp\left[\frac{e^{-\left(\frac{x}{\sigma}\right)^\mu} - 1}{\beta}\right]}{\left(e^{\left(\frac{x}{\sigma}\right)^\mu} - 1\right) \left[\Gamma\left(\alpha, \frac{1}{\beta}\right) - \Gamma\left(\alpha, \frac{1-e^{-\left(\frac{x}{\sigma}\right)^\mu}}{\beta}\right)\right]}.$$

For some values of the parameters, the plots of the density and the hazard rate function are shown in Figures 1 and 2, respectively. It can be summarized some of the shape properties of the GW distribution as follows:

- The distribution is right-skewed when $\alpha > 0.5$. As σ increases the degree of right skewness increases.
- The distribution is reverse "J" shaped when $\mu < 1, \alpha < 1$, also $\mu > 1, \alpha < 0.5$.

It is depicted by Figure 2 that the hazard rate function of the GW distribution can take monotonic, bathtub and unimodal-bathtub shapes for different parametric combinations.

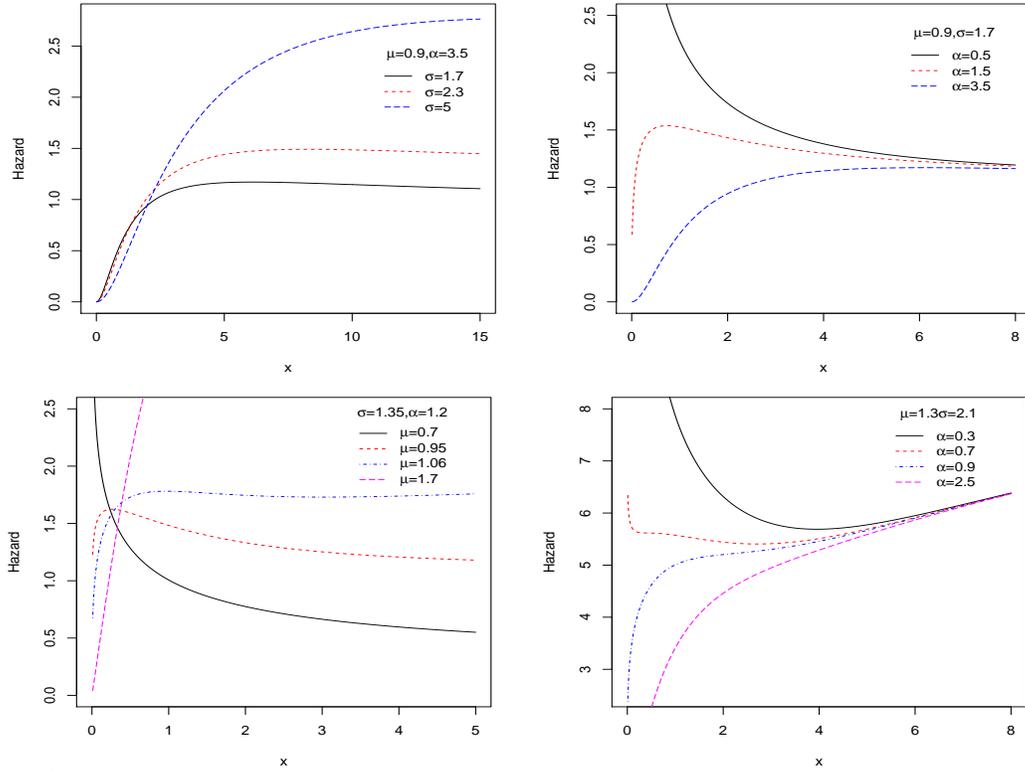


Figure 2: Plots of the hazard rate function of the $GW(\alpha, 1, \mu, \sigma)$ distribution for some various values of parameters.

3 Moments

In general, k -th non central moment of a GW distribution cannot be easily evaluated. considering common definition, k -th non central moment can be written as following

$$E[X^k] = \frac{\mu x^{\mu+k-1} \left(\frac{1-e^{-\left(\frac{x}{\sigma}\right)^\mu}}{\beta} \right)^\alpha \exp \left\{ \frac{e^{-\left(\frac{x}{\sigma}\right)^\mu} - 1}{\beta} \right\}}{\sigma^\mu \left(e^{\left(\frac{x}{\sigma}\right)^\mu} - 1 \right) \left[\Gamma(\alpha) - \Gamma \left(\alpha, \frac{1}{\beta} \right) \right]}.$$

Hence, these measures for the $GW(\alpha, \beta, 0.5, 0.7)$ distribution are calculated and presented in Table 1 for various combinations of α and β . From Table 1, it can be concluded that for fixed α , the mean, the

Table 1: Mean, second moments, variance, skewness and kurtosis when $\mu = 0.5$ and $\sigma = 0.7$ for various values of α and β .

α	β	Mean	$E[X^2]$	$\text{Var}(X)$	Skewness	Kurtosis
0.5	0.1	0.007	0.002	0.002	197.803	144131.7
	0.5	0.245	1.459	1.399	16.608	545.097
	1	0.439	3.053	2.861	11.933	276.483
	2	0.581	4.315	3.977	10.228	201.946
1	0.1	0.022	0.012	0.011	106.469	36134.95
	0.5	0.592	3.947	3.596	10.556	218.449
	1	0.934	7.073	6.201	8.205	130.650
	2	1.152	9.211	7.884	7.333	104.097
3	0.1	0.189	0.407	0.371	26.021	1595.047
	0.5	2.238	19.597	14.589	5.458	58.184
	1	2.762	26.047	18.419	4.902	46.962
	2	3.033	29.558	20.361	4.680	42.851

second moments and the variance are increasing functions of β , while the skewness and the kurtosis is a decreasing function of β . Also, for fixed β , the mean, the second moments and the variance are increasing functions of α , while the skewness and the kurtosis is a decreasing function of α . Table 1 also shows that the GW distribution is right skewed. Over-dispersion in a distribution is a situation in which the variance exceeds the mean, under-dispersion is the opposite, and equi-dispersion occurs when the variance is equal to the mean. From Table 1, the GW distribution satisfies the over-dispersion property almost for all values of parameters.

4 Parameter estimation and inference

Let X_1, \dots, X_n be a random sample, with observed values x_1, \dots, x_n from $GW(\alpha, \beta, \mu, \sigma)$. The log-likelihood function for the vector of parameters $\Theta = (\alpha, \beta, \mu, \sigma)^T$ can be written as

$$\begin{aligned}
 l_n = l_n(\Theta) &= \alpha \sum_{i=1}^n \log \left(e^{-\left(\frac{x_i}{\sigma}\right)^\mu} - 1 \right) + \sum_{i=1}^n \left(e^{-\left(\frac{x_i}{\sigma}\right)^\mu} - 1 \right) - \sum_{i=1}^n \log \left(e^{-\left(\frac{x_i}{\sigma}\right)^\mu} - 1 \right) \\
 &+ \mu \sum_{i=1}^n \log \left(\frac{x_i}{\sigma} \right) - \sum_{i=1}^n \log(x_i) - n \log \left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{1}{\beta}\right) \right] + n \log(\mu)
 \end{aligned}$$

The components of the score vector $U(\Theta) = \left(\frac{\partial l_n}{\partial \alpha}, \frac{\partial l_n}{\partial \beta}, \frac{\partial l_n}{\partial \mu}, \frac{\partial l_n}{\partial \sigma} \right)^T$ are given by

$$\begin{aligned} \frac{\partial l_n}{\partial \alpha} &= \sum_{i=1}^n \log \left(e^{-\left(\frac{x_i}{\sigma}\right)^\mu} - 1 \right) - \frac{n \psi(\alpha) \Gamma(\alpha) - n G_{2,3}^{3,0} \left(\frac{1}{\beta} \middle| \begin{matrix} 1, 1 \\ 0, 0, \alpha \end{matrix} \right) - n \log \left(\frac{1}{\beta} \right) \Gamma \left(\alpha, \frac{1}{\beta} \right)}{\Gamma(\alpha) - \Gamma \left(\alpha, \frac{1}{\beta} \right)}, \\ \frac{\partial l_n}{\partial \beta} &= \frac{e^{-1/\beta} n \left(\frac{1}{\beta} \right)^{\alpha+1}}{\Gamma(\alpha) - \Gamma \left(\alpha, \frac{1}{\beta} \right)}, \\ \frac{\partial l_n}{\partial \mu} &= \frac{n}{\mu} + (\alpha - 1) \sum_{i=1}^n \frac{e^{-\left(\frac{x_i}{\sigma}\right)^\mu} \left(\frac{x_i}{\sigma}\right)^\mu \log \left(\frac{x_i}{\sigma}\right)}{e^{-\left(\frac{x_i}{\sigma}\right)^\mu} - 1} + \sum_{i=1}^n -e^{-\left(\frac{x_i}{\sigma}\right)^\mu} \left(\frac{x_i}{\sigma}\right)^\mu \log \left(\frac{x_i}{\sigma}\right) + \sum_{i=1}^n \log \left(\frac{x_i}{\sigma}\right), \\ \frac{\partial l_n}{\partial \sigma} &= (\alpha - 1) \sum_{i=1}^n \frac{\mu x_i e^{-\left(\frac{x_i}{\sigma}\right)^\mu} \left(\frac{x_i}{\sigma}\right)^{\mu-1}}{\sigma^2 \left(e^{-\left(\frac{x_i}{\sigma}\right)^\mu} - 1 \right)} + \sum_{i=1}^n \frac{\mu x_i e^{-\left(\frac{x_i}{\sigma}\right)^\mu} \left(\frac{x_i}{\sigma}\right)^{\mu-1}}{\sigma^2} - \frac{\mu n}{\sigma}, \end{aligned}$$

where $\psi(\cdot)$ is the digamma function and $G_{p,q}^{m,n}(\cdot)$ is the Meijer G-function, which digamma function is the logarithmic derivative of the gamma function and Meijer G-function is defined by

$$G_{m,n}^{p,q} \left(z \middle| \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right) = \frac{1}{2\pi i} \int_{\gamma_L} \frac{\prod_{j=1}^m \Gamma(b_j - s) \prod_{j=1}^n \Gamma(1 - a_j + s)}{\prod_{j=n+1}^p \Gamma(a_j - s) \prod_{j=m+1}^q \Gamma(1 - b_j + s)} z^s ds,$$

where the contour γ_L lies between the poles of $\Gamma(1 - a_i - s)$ and the poles of $\Gamma(b_i + s)$. The maximum likelihood estimation (MLE) of Θ , say $\hat{\Theta}$, is obtained by solving the nonlinear system $U(\Theta) = \mathbf{0}$. This nonlinear system of equations does not have a closed form.

Under conditions that are fulfilled for parameters in the interior of the parameter space but not on the boundary, the asymptotic distribution of $\sqrt{n}(\hat{\Theta} - \Theta)$ is $N_4(\mathbf{0}, I^{-1}(\Theta))$ where $I^{-1}(\Theta)$ is the inverse of the Fisher information matrix. An $100(1 - \gamma)$ asymptotic confidence interval (ACI) for each parameter Θ_i is given by

$$AIC(\Theta_i) = \left(\hat{\Theta}_i - z_{\alpha/2} \sqrt{\hat{I}^{-1}(\hat{\Theta})_i/n}, \hat{\Theta}_i + z_{\alpha/2} \sqrt{\hat{I}^{-1}(\hat{\Theta})_i/n} \right),$$

where $\hat{I}^{-1}(\hat{\Theta})_i$ is the i -th diagonal element of $\hat{I}^{-1}(\hat{\Theta})$.

5 Applications

In this section, the flexibility and applicability of the proposed model is illustrated as compared to the alternative gamma-Weibull (GW.a) introduced by [Alzaatreh et al. \(2014\)](#) and gamma-Weibull (GW.p) presented by [Provost et al. \(2011\)](#) distributions. The gamma-Weibull model is applied to a data set published in [Suprawhardana et al. \(1999\)](#) which consists of time between failures (thousands of hours) of secondary reactor pumps. The data set consists of 23 observations. The TTT plot of this set of data in [Figure 3](#) displays a bathtub-shaped hazard rate function that indicates the appropriateness of the GW distribution to fit these data.

In order to compare the models, the MLEs of the parameters, $-2\log$ -likelihood, the Kolmogorov-Smirnov test statistic (K-S), P-value, the Anderson-Darling test statistic (AD), the Cramér-von Mises test statistic (CM) and Durbin-Watson test statistic (DW) are given in [Table 2](#) for this data set. The CM and DW test statistics are described in details in [Chen and Balakrishnan \(1995\)](#) and [Watson \(1961\)](#), respectively. In general, the smaller the values of K-S, AD, CM and WA, the better the fit to the data. From the values of these statistics, we conclude that the GW distribution provides a better fit to this data than the other models.

The fitted densities and the empirical distribution versus the fitted cumulative distribution functions of NP, N and SN models are displayed in [Figure 4](#). These plots suggest that the GW distribution is superior to the other distributions in terms of model fitting.

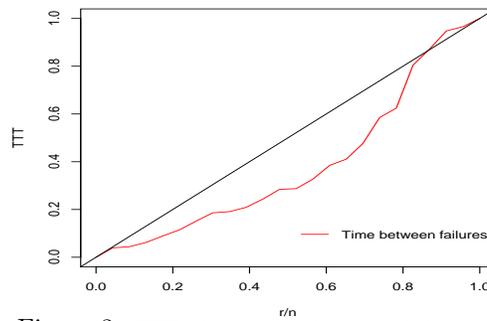


Figure 3: TTT plot to time between failures data

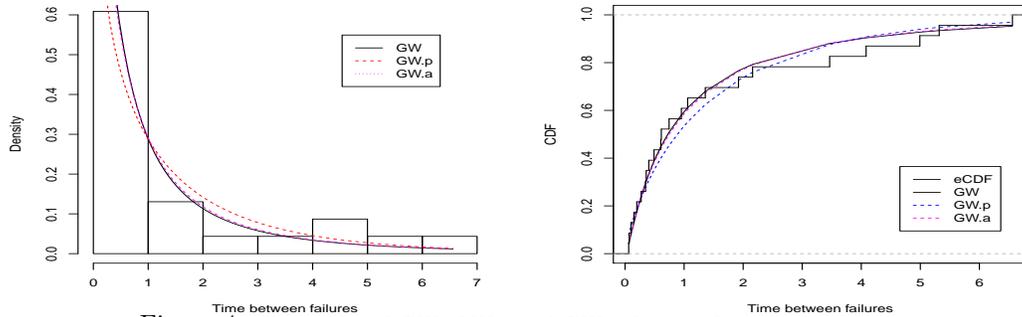


Figure 4: Plots of fitted GW, GW.a and GW.p for time between failures data.

Table 2: MLEs, -2 Log L, K -S, P-value, AD, CM and DW statistics for time between failures data.

Dist	MLE	-2 Log L	K-S	P-value	AD	CM	DW
GW	$\hat{\alpha} = 8.264, \hat{\mu} = 0.333,$ $\hat{\sigma} = 0.051$	63.639	0.098	0.97	0.233	0.107	0.107
GW.a	$\hat{\alpha} = 16.197, \hat{\gamma} = 0.054,$ $\hat{c} = 0.184, \hat{\beta} = 0.102$	63.635	0.099	0.96	0.246	0.110	0.109
GW.p	$\hat{k} = 0.814, \hat{\xi} = 0.010,$ $\hat{\lambda} = 1.419$	65.045	0.119	0.87	0.406	0.142	0.138

References

- Alzaatreh, A., Famoyeh, F. and Lee, C. (2014). The gamma-normal distribution: Properties and applications, *Computational Statistics and Data Analysis*, **69**, 67–80.
- Chen, G. and Balakrishnan, N. (1995), A general purpose approximate goodness-of-fit test, *Journal of Quality Technology*, **27**, 154–161.
- Johnson, N. and Kotz, S. (1976). *Distributions in Statistics, Continuous Univariate Distributions-1*, Wiley, New York.
- Meeker, W. and Escobar, L. (1988). *Statistical Methods for Reliability Data*. Wiley, New York.
- Nadarajah, S. and Kotz, S. (2006). On the product and ratio of gamma and Weibull random variables, *Econometric Theory*, **22**, 338–344.

- Prentice, R. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika*, **61**, 539–544.
- Provost, S.B., Saboor, A. and Ahmad, M. (2011). The gamma-normal distribution, *Pakistan Journal of Statistic*, **27(2)**, 111–131
- Smith, R. and Naylor, R. (1987). A comparison of maximum likelihood and Bayesian estimators for the three parameter Weibull distribution. *Applied Statistics*, **36**, 358–369.
- Suprawhardana, S. Prayoto, M, and Sangadji. (1999). Total time on test plot analysis for mechanical components of the RSG-GAS reactor, *Atom Indones*, **25(2)**.
- Watson, G.S. (1961). Goodness-of-fit tests on a circle, *Biometrika*, **48**, 109–114.



The Cowan's Copula and its Extension

S.M. Mirhosseini¹, A. Dolati²

¹mmirhoseini@yazd.ac.ir

²adolati@yazd.ac.ir

Abstract: This paper considers a class of absolutely continuous bivariate exponential distributions whose univariate margins are the ordinary exponential distributions. We study different mathematical properties of the Cowan's model. We introduce some extensions and study their properties. **Keywords** Bivariate exponential distribution; Copula; Dependence; Measure of association;

Mathematics Subject Classification (2010): Primary 62E15; Secondary 62H10.

1 Introduction

In many reliability situations bivariate lifetime data arise and in these cases it is important to consider different distributions that could be used to model such data. The exponential distributions are perhaps the most widely applied statistical distributions in reliability and survival analysis. A large number of bivariate exponential distributions have been proposed in literature, see, e.g., Arnold et al. (1988), Block et al. (1974), Dimitrakopoulou et al. (2012); Downton (1970); Freund (1961); Gumbel (1960); Hawkes (1972); Marshall et al. (1967b). A wide survey on bivariate distributions could be seen in Balakrishnan et al. (2009); Kotz et al. (1988). Cowan (1987) introduced the new bivariate exponential distribution which arises in random geometry with marginal ordinal exponential distributions. In this paper, starting from Cowan's bivariate distribution, we study some properties in view copula. In the following, we present an extension of Cowan's copula and investigate its dependence structure.

Let $F(x, y)$ be a Cowan's bivariate distribution with univariate marginal exponential mean parameter λ .

¹Speaker's Name : S. M. Mirhosseini

Then we have

$$F_\theta(x, y) = 1 - e^{-\lambda x} - e^{-\lambda y} + e^{-\frac{\lambda}{2}(x+y+\sqrt{(x+y)^2-2\theta xy})}, \quad x, y \geq 0 \quad 0 \leq \theta \leq 1. \tag{1.1}$$

where $\theta = \frac{1+\cos a}{2}$ and a is angle between two random line which generate from poisson process with rate of λ . The survival function of Cowan is given by

$$\bar{F}_\theta(x, y) = e^{-\frac{\lambda}{2}(x+y+\sqrt{(x+y)^2-4\theta xy})}, \quad x, y \geq 0 \quad 0 \leq \theta \leq 1. \tag{1.2}$$

Let $C(u, v)$ and $\bar{C}(u, v)$ are copula and survival copula, respectively. Then, the corresponding copulas, the survival of Cowan's copula and Cowan's copula, are given by

$$\bar{C}_\theta(u, v) = \sqrt{uv}e^{-\frac{1}{2}\sqrt{(\ln(uv))^2-4\theta \ln u \ln v}}, \quad u, v \in [0, 1]. \tag{1.3}$$

$$\begin{aligned} C_\theta(u, v) &= u + v - 1 + \bar{C}_\theta(1 - u, 1 - v) \\ &= u + v - 1 + \sqrt{(1 - u)(1 - v)}e^{-\frac{1}{2}\sqrt{(\ln(1-u)(1-v))^2-4\theta \ln(1-u) \ln(1-v)}}, \quad u, v \in [0, 1]. \end{aligned} \tag{1.4}$$

Noting that the simplicity form of (1.3), the following we investigate the survival Cowan's Copula.

Proposition 1.1. *Let $\bar{C}(u, v)$ be a survival Cowan's copula, then Cowan's copula family is positively ordered with respect to parameter of θ . i.e. for all $\theta_1 \leq \theta_2$ we have $\bar{C}_{\theta_1}(u, v) \leq \bar{C}_{\theta_2}(u, v)$.*

Proof. For $\theta_1 \leq \theta_2$, we have

$$\frac{\bar{C}_{\theta_1}(u, v)}{\bar{C}_{\theta_2}(u, v)} = e^{-\frac{A}{2}}$$

where $A = \frac{4(\theta_2 - \theta_1) \ln u \ln v}{\sqrt{(\ln(uv))^2 - 4\theta_1 \ln u \ln v} + \sqrt{(\ln(uv))^2 - 4\theta_2 \ln u \ln v}} > 0$. Then

$$\frac{\bar{C}_{\theta_1}(u, v)}{\bar{C}_{\theta_2}(u, v)} \leq 1.$$

The proof is complete. □

It is obvious that $uv = \bar{C}_{\theta=0}(u, v) \leq \bar{C}_\theta(u, v) \leq \bar{C}_{\theta=1}(u, v) = \min(u, v)$.

Corollary 1.2. *Let $\bar{C}(u, v)$ be a survival Cowan's copula, then $\bar{C}(u, v)$ is positively quadrant dependent (PQD).*

Proposition 1.3. For $\bar{C}(u, v)$ given (1.3), we have

$$\lambda_l = 0, \quad \lambda_u = 1 - \sqrt{1 - \theta} \text{ and } \beta = 2^{1 + \sqrt{1 - \theta}} - 1.$$

Corollary 1.4. For $\bar{C}(u, v)$ given in (1.3), we have

$$E_l = \frac{\lambda_l}{2 - \lambda_l} = 0 \text{ and } E_u = \frac{\lambda_u}{2 - \lambda_u} = \frac{1 - \sqrt{1 - \theta}}{1 + \sqrt{1 - \theta}}$$

Proposition 1.5. Let $\bar{F}(x, y)$ be a survival distribution function given in (1.2), then X and Y are RCSI.

Proof. By simple calculating, we drive $\frac{\partial^2}{\partial x \partial y} \log \bar{F}(x, y) = \frac{2\lambda\theta xy(1-\theta)}{\{(x+y)^2 - 4\theta xy\}^{\frac{3}{2}}}$. Since $x, y \in [0, 1]$, $\lambda > 0$ and $0 \leq \theta \leq 1$, we can conclude that $\frac{\partial^2}{\partial x \partial y} \log \bar{F}(x, y) > 0$. The proof is complete. \square

Remark 1.6. Using proposition 1.5 and denoting Oakes's local dependence measure [Holland et al. \(1987\)](#), we find out that $\theta(x, y) > 1$ and $\lambda(x, y) > 1$. Therefore, we can show that $r(x, y) \geq r_1(x, y)r_2(x, y)$, where $r(x, y) = \frac{f(x, y)}{F(x, y)}$ is the bivariate hazard rate which is introduced by Basu and $r_i(x_1, x_2) = -\frac{\partial \log \bar{F}(x_1, x_2)}{\partial x_i}$, $i = 1, 2$.

Remark 1.7. The survival of Cowan's copula, $\bar{C}(u, v)$, could not be Archimedean copula, Because

$$0.008612 = \bar{C}(0.5, \bar{C}(0.25, 0.05)) \neq \bar{C}(\bar{C}(0.5, 0.25), 0.05) = 0.008853$$

for fixed $\theta = 0.2$.

Proposition 1.8. Suppose $\bar{C}(u, v)$ be a survival Cowan's copula given in (1.3). Then $\bar{C}(u, v)$ is invariant under distortion function $g(t) = t^\gamma$, where $0 \leq \gamma \leq 1$.

Proof. For poof, we must show that

$$\bar{C}_g(u, v) = g(\bar{C}(g^{-1}(u), g^{-1}(v))) = \bar{C}(u, v).$$

By straightforward calculation, we drive it. \square

Remark 1.9. More, if $\gamma = \frac{1}{n}$, then $\bar{C}^n(u, v) = \bar{C}(u^n, v^n)$. we can conclude that $\bar{C}(u, v)$ is Extreme Value (EV) copula.

2 The extension of Cowan's copula

In this section, we introduce an extension of Cowan's copula which have an new parameter.

Proposition 2.1. For $0 \leq \theta \leq 1$ and $0 \leq \alpha \leq 1$, $C_\alpha(u, v)$ given by

$$C_\alpha(u, v) = \Pi(u^\alpha, v^\alpha) e^{-(1-\alpha)\sqrt{(\ln uv)^2 - 4\theta \ln u \ln v}} \quad (2.1)$$

is a copula function.

Proof. It is easy to check that $C_\alpha(u, 0) = C_\alpha(0, v) = 0$, $C_\alpha(u, 1) = u$ and $C_\alpha(1, v) = v$. For proof of 2-Increasing property, we can show that $\frac{\partial^2}{\partial u \partial v} C_\alpha(u, v) > 0$. Thus, this is equivalent to $\frac{\partial^2}{\partial u \partial v} \log C_\alpha(u, v) > 0$. By some easy calculating, we drive

$$\frac{\partial^2}{\partial u \partial v} \log C_\alpha(u, v) = \frac{4\theta(1-\theta)(1-\alpha) \ln u \ln v}{\{(\ln uv)^2 - 4\theta \ln u \ln v\}^{3/2}}.$$

Denoting $u, v \in [0, 1]$, $0 \leq \theta \leq 1$ and $0 \leq \alpha \leq 1$, what is need to prove, we received. \square

It is clear that if $\alpha = \frac{1}{2}$, Cowan's copula is obtained. Following , we probe about positively or negatively ordered for this extension.

Proposition 2.2. Let $C_\alpha(u, v)$ be a copula given in (2.1), then proposed copula family is positively ordered with respect to parameter of θ and negatively ordered with respect to parameter of α .

It is easy to see that $uv = C_{\alpha=1}(u, v) \leq C_\alpha(u, v) \leq C_{\alpha=\frac{1}{2}}(u, v) = \text{Cowan's copula}$ for $\alpha > \frac{1}{2}$. Therefore, this copula is PQD show that the copula (2.1) is ne

Conclusion

In this paper, we studied about Cowan copula : Cowan's copula family is positively ordered with respect to parameter of θ and it is extreme value (EV) copula.

References

- Arnold, B. C. and Strauss, D. (1988). Bivariate distributions with exponential conditionals. *J. Amer. Statist. Assoc.*, **83**, 522–527.

- Balakrishnan, N. and Lai, C. D. (2009). *Continuous Bivariate Distributions*. second edition. Springer, New York.
- Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential extension. *J. Amer. Statist. Assoc.*, **69**, 1031–1036.
- Cowan, R. (1987). A bivariate exponential distribution arising in random geometry. *Annals of the Institute of Statistical Mathematics*, **39**, 103–111.
- Dimitrakopoulou, T., Adamidis, K. and Loukas, S. (2012). Bivariate extended exponential-geometric distributions. *Communications in Statistics-Theory and Methods*, **41**, 1129–1150.
- Downton, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society*, **32 (3)**, 408–417.
- Freund, J. (1961). A bivariate extension of the exponential distribution. *J. Amer. Statist. Assoc.*, **56**, 971–977.
- Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, **55**, 698–707.
- Hawkes, A. G. (1972). A bivariate exponential distribution with applications to reliability. *Journal of the Royal Statistical Society, B*, **34**, 129–131.
- Holland, P. W. and Wang, Y. J. (1987). Dependence functions for continuous bivariate densities. *Commun. Statist. Theory and Methods*, **16:3**, 863–876.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.
- Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). *Continuous multivariate distributions*, Vol. 1, second ed., Wiley: New York.
- Nelsen, R.B. (2006). *An Introduction to Copulas*. Second Edition, Springer, New York.
- Marshall, A. W. and Olkin, I. A. (1967b). A generalized bivariate exponential distribution, *J. Appl. Prob.*, **4 (2)**, 291–302.

Pulcini, G. (2006). A Bivariate Distribution for the Reliability Analysis of Failure Data in Presence of a Forewarning or Primer Event. *Communications in Statistics Theory and Methods*, **35**, 2107–2126.

Shaked, M. & Shantikumar, G. (2007), *Stochastic Orders*, Springer, New York.



Longitude Shape Analysis by Using the Spherical Coordinates

Meisam Moghimbeygi, Mousa Golalizadeh¹

Department of Statistics, Tarbiat Modares University, Tehran, Iran

Abstract: The longitudinal shape analysis is one of the important topics in morphometry. Up to the Kendall's definition of shape, the shape appertains of non-Euclidean space, so longitudinal study of shape on sphere is pursued in this paper. Using the triangulation and size of shape, the prediction of the shape of the configuration in a period of time will be considered. Also, real data analysis will be discussed and the relevant interpretations are presented.

Keywords Longitudinal Model, Nonparametric Inference, Shape Analysis, Triangulation, Non-Euclidean space.

Mathematics Subject Classification (2010): 62G08.

1 Introduction

Statistical analysis of shape became popular in the late 1970 and early 1980s by a brief note of [Kendall \(1977\)](#) on “diffusion of shape”. He then completed his theory ([Kendall, 1984](#)) and provided further idea on his discussion on the paper written by [Bookstein \(1986\)](#). Hence, it can be claimed that modern extension of this field of statistics is own to late Kendall and Bookstein.

The longitudinal analysis of the shape is of great interest in morphometry. Much research has been conducted in the field of longitudinal shape analysis with the view to morphometrics. However, they often did not have a proper statistical analysis in their merits. For example, [Buschang et al. \(1993\)](#) studied the growth of nasal in persons and longitudinal shape change of the dorsum. They analyzed the changes of angles by rotation of shape to lie at a fix point. [Sharma et al. \(2014\)](#) discuss the age changes of the hard and soft tissues of human face.

¹Mousa Golalizadeh: golalizadeh@modares.ac.ir

An important problems in statistical shape analysis is to develop some methods for studying the shape changes over time. One of the common procedures in this respect is to approximate the non-Euclidean space of shape with its linear counterpart. This view has been undertaken in many literatures. For instance, [Morris et al. \(1999\)](#) analysed the human face change during their ages. They presented a semi-automatic method to perform Procrustes analysis and principal component analysis on shape data. [Le and Kume \(2000\)](#) studied changes of shape mean over time with the use of classical multidimensional scaling. They compared the shape changes using the first principal coordinates of mean shapes. [Kent et al. \(2000\)](#) fitted the growth model for landmark data using the Procrustes tangent coordinates where the objects lie in ambient space.

Because the shape space is hyper-sphere, to propose the longitudinal shape models can be studied through invoking such models on general sphere, from topological view point. In particular, considering the Bookstein shape coordinate systems, which are based upon triangulation ([Bookstein , 1986](#)), the shapes can be represented by points on sphere as [Kendall \(1984\)](#) has shown. The important remark on selecting two far points, as the base lines, are the only criterion in this methodology.

There are many methods to fit a curve to spherical data. [Gould \(1969\)](#) used the angles of points on sphere to introduce two separate regression models in which the Fisher distribution is considered as the density for errors. Following [Gould \(1969\)](#), [Thompson and Clark \(1981\)](#) proposed non-parametric version of two separated models for angles by using cubic spline and least square methods. Their methods fail in the vicinity of poles and so the data must keep away from the pole as far as possible. [Clark and Thompson \(1984\)](#) overcame this obstacle via using the tangent plane on the pole and proposed fitting the spline function for any coordinate. [Parker and Denham \(1979\)](#) introduced spline function without restricting data to be on the sphere. Although this method seems to be useful for the concentrated data, it fails for the well-scattered objects. [Fisher et al. \(1987\)](#) proposed two families of the spherical spline on the line. They introduced two families of curves by using the relevant differential geometry that are suitable for spline fitting and to be invariant under the choice of coordinates system.

It is known from elementary statistics that one of the methods to predict the responses is nonparametric smoothing. The study of smoothing models on a circle were considered in many literatures. [Prentice \(1987\)](#) fitted the spline path to circular data by using the rotation parameters. [Di Marzio](#)

et al. (2013) presented methods to fit the nonparametric spline to circular data by minimizing the suitable risk function. In this paper, we use nonparametric smoothing procedure on a circle and extend it to the sphere with minimizing the Euclidean risk function.

Main aim of this paper is to study of shape change in a period of time using nonparametric models on a sphere. Section 2, describes longitudinal model for spherical data. General idea to have shape growth model is given in Section 3. The application of the model on real data is presented in Section 4. General results and relevant discussions are also presented.

2 Longitudinal Model for Spherical Data

We first assume that the observations on sphere have independent angles. Gould (1969) considered the independent variables on sphere through using linear type models. Utilizing his idea, Thompson and Clark (1981) introduced the parametric models for angles near the pole with Fisher distribution. Also, using the cubic splines they presented non-parametric models for spherical data identified by angles. Their models are well described by two measurable functions, $f : t \rightarrow \phi$ and $g : t \rightarrow \theta$, for $i = 1, \dots, n$ written as

$$\begin{cases} \Phi_i = f(t_i) + \varepsilon_{i1} \\ \Theta_i = g(t_i) + \varepsilon_{i2} \end{cases} \quad (2.1)$$

where ε_{i1} and ε_{i2} are the errors of model.

Thompson and Clark (1981) also introduced parametric model based on an approximation of the Fisher distribution. They supposed that θ_i 's are normally distributed with mean $g(t_i)$ and variance $1/\kappa$. Furthermore, it is assumed that the angles ϕ 's are normally distributed with mean $f(t_i)$ and variance $1/\kappa \sin^2 g(t_i)$. For this setting, the angles θ_i and ϕ_i are approximately independent. Di Marzio et al. (2013) introduce a non-parametric model for angular observation ψ_i at time t_i as $\psi_i = h(t_i) + \varepsilon_i$ where ε_i is a random variable with zero mean direction and finite concentration. In order to estimate $h(\cdot)$ in some sensible way, they considered a proper risk function. To propose a non-parametric model on sphere, we basically extend Di Marzio et al. (2013) methods on circle to the sphere. In fact, we combine their ideas with the procedure described by Thompson and Clark (1981)

to derive a proper model to tackle the longitudinal analysis of the spherical data. We first proposed the following Lemma.

Lemma 2.1. For $t \in T$, let $m_1(t) = E[\sin \theta \sin \phi | T = t]$, $m_2(t) = E[\sin \theta \cos \phi | T = t]$, $m_3(t) = E[\sin \theta \cos(\phi - f_t) | T = t]$ and $m_4(t) = E[\cos \theta | T = t]$. Then, the risk function

$$R[(\theta, \phi), (f_t, g_t)] = E[2 - 2 \sin \theta \sin g_t \cos(\phi - f_t) - 2 \cos \theta \cos g_t] \quad (2.2)$$

is minimized when $f_t = \text{atan2}(m_1(t), m_2(t))$ and $g_t = \text{atan2}(m_3(t), m_4(t))$, where the function $\text{atan2}(y, x)$ refer to the angle between the x -axis and the vector (x, y) .

Proof. For minimizing the risk function, we take derivatives of the expectation (2.2) with respect to f and g . They are as follow

$$\begin{aligned} & \partial E[2 - 2 \sin \theta \sin g_t \cos(\phi - f_t) - 2 \cos \theta \cos g_t | t] / \partial f \\ &= 2E[\sin \theta \cos \phi | t] \sin g_t \sin f_t - 2E[\sin \theta \sin \phi | t] \sin g_t \cos f_t \\ &= m_2(t) \sin g_t \sin f_t - m_1(t) \sin g_t \cos f_t = 0, \end{aligned}$$

and

$$\begin{aligned} & \partial E[2 - 2 \sin \theta \sin g_t \cos(\phi - f_t) - 2 \cos \theta \cos g_t | t] / \partial g \\ &= -2 \cos g_t E[\sin \theta \cos(\phi - f_t) | t] + 2 \sin g_t E[\cos \theta | t] \\ &= m_3(t) \cos g_t + m_4(t) \sin g_t = 0. \end{aligned}$$

Because the risk function is convex, then it is clearly minimized when $f_t = \text{atan2}(m_1(t), m_2(t))$ and $g_t = \text{atan2}(m_3(t), m_4(t))$ and these will prove the assertion. \square

To introduce a paper smoother, we define the sample statistics as

$$\begin{aligned} \hat{m}_1(t) &= \frac{1}{n} \sum_{i=1}^n \sin \Theta_i \sin \Phi_i K_{h_1}(T_i - t) \\ \hat{m}_2(t) &= \frac{1}{n} \sum_{i=1}^n \sin \Theta_i \cos \Phi_i K_{h_1}(T_i - t) \\ \hat{m}_3(t) &= \frac{1}{n} \sum_{i=1}^n \sin \Theta_i \cos(\Phi_i - \hat{f}_t) K_{h_2}(T_i - t) \\ \hat{m}_4(t) &= \frac{1}{n} \sum_{i=1}^n \cos \Theta_i K_{h_2}(T_i - t) \end{aligned}$$

where K_{h_1} and K_{h_2} are kernel weight and

$$\hat{f}(t) = \text{atan2}(\hat{m}_1(t), \hat{m}_2(t)), \quad \hat{g}(t) = \text{atan2}(\hat{m}_3(t), \hat{m}_4(t)). \quad (2.3)$$

Note that the estimator of f is independent of g , while the estimator of g depends on the estimate of f . In order to select an appropriate kernel function, we can use a density function having its maximum at 0.

3 Shape Growth Model

We consider the triangular shape space which is simply the unit sphere with radius $1/2$ [Kendall \(1984\)](#). Hence, for modeling the shape we need to look at triangulation of shape. Because, we lose all of information about size of configuration using the Kendall shape coordinate, then, we can use the size of configuration after reconstruction of shape for modeling the growth. In this section, we survey the procedure for longitudinal triangulation of shapes and then describe the growth models for configurations.

Since our approach is based on triangulation of objects, a suitable method of triangulation is necessary to implement. Therefore, we must choose a method providing the least error on reconstructing the shapes. It seems that using the independent triangles will lead to the less cumulative errors. It means, fixing two appropriate landmarks as a baseline for all triangles and using the coordinates of the remaining landmark will help us to achieve the best performance. As we know, choosing the baseline is one of the important issues in shape statistics and morphometry. [Bookstein \(1991\)](#) showed that the selection of baseline will be much more effective to view the triangles after translation, rotation and rescaling. Generally, the best way is to select baselines as the two furthest landmarks that passes through the centroid of the objects. As pointed out by [Zelditch et al. \(2012\)](#), the interpretation of shapes with the longest baseline is much more easier than having them short in length.

It is obvious that changing the length of baselines leads to a change in shape. Therefore, we pay attention to the distance between landmarks considered as baselines. Then, constructing the regression model can be modeled based upon the change of the baselines as the response variable versus time. Having said that, the configuration would indirectly play their roles in the regression model through triangulation where the baseline will always be one of sides of triangle.

4 Application

In this section we are interested on employing the proposed model on real data. So, we choose the data from Bookstein (1991) dealing with laboratory rats from a close bred European. The data include 18 male laboratory rats in 8 ages and 8 landmarks. A schematic plot of data is shown in Figure 1 in biological significant: basion, opisthion, interparietal suture, lambda, bregma, speno-ethmoid synchondrosis, intersphenoidal suture and speno-occipital synchondrosis. The mean shapes is also plotted on the right hand side of Figure 1.

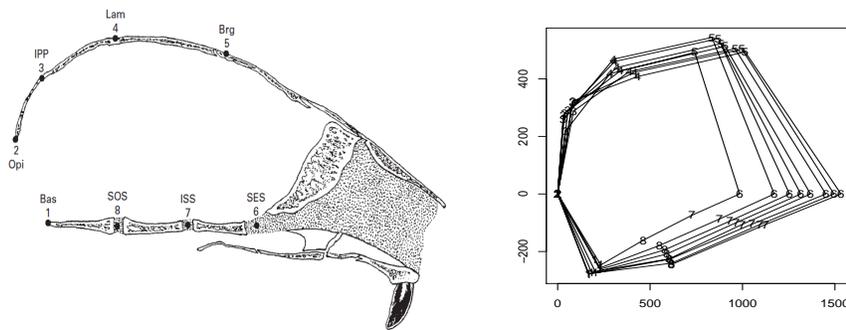


Figure 1: Left to right: the location of the landmarks to display rat skulls that derived from Bookstein (1991) and mean shape of 18 rat skull in 8 times.

At the first step, we calculate the Bookstein shape coordinate of rat configuration. The landmarks numbered 2 and 6 are the furthest points away and so they are considered as baseline points. To invoke equations (2.3), we selected W_1 and W_2 as weights coming from centered normal distribution with variance 6 for both. Then, we employed the longitudinal model for triangulation rat shapes. The prediction of models can be seen in the left hand side of Figure 2 for 150 consecutive days as solid lines. The sum of square error in our model is 6×10^{-4} , the R-square for models are 0.99, NA, 0.95, 0.99, 0.98, NA, 0.93, 0.99, respectively and the total R-square is 0.98.

At the second step, we are interested on the size of shapes. To this end, we used the size of the baseline and the least square method to find the best model for the sizes. Since there are many models to fit the data, we utilized the curve fitting toolbox in MATLAB to find the best model from proposed

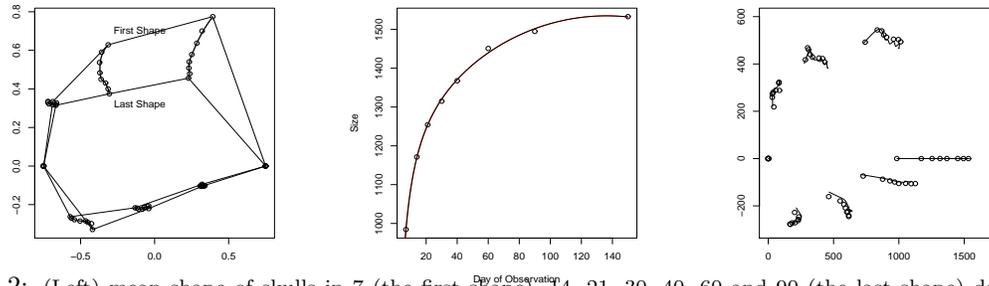


Figure 2: (Left) mean shape of skulls in 7 (the first shape), 14, 21, 30 ,40, 60 and 90 (the last shape) days and its prediction shown by the solid line in 7-150 days. (Middle) plot of size versus time that solid line is in fact its prediction. (Right) mean of configurations (points) with its predictions (solid lines).

defaults. One of them for our data is rational model given as

$$s_i = \frac{a_1 + a_2t_i + a_3t_i^2 + a_4t_i^3}{t_i + b} + \varepsilon_i, \quad i = 1, \dots, 8. \tag{4.1}$$

Hence, the R-square is 0.9991 with the fitted model shown at the middle hand side of Figure 2.

At the last step, we combined the size of the shape and the longitudinal shape model to describe the growth. This result is shown at right hand side of Figure ???. The R-square criteria for 8 models are 0.43, *Na*, 0.92, 0.86, 0.96, 0.99, 0.95, 0.84, respectively and the total R-square is 0.95. The general results show that our procedures did manage to represent proper model for longitudinal variation at shape and size.

Conclusion

In analyzing longitudinal shape data, the parameter estimation gets serious problem by increasing the number of landmarks. One way to overcome this problem is to reduce the data dimension using triangulation of shape. Triangle as the most primitive shape have a key role in the statistical analysis of shapes. We dealt with a significant dimension reduction by providing appropriate triangulation and using them to provide a longitudinal shape model. Since the shape space is sphere, it seems that extending the statistical analysis in this space is much more easier than complex projective space. The difficulty arises when one pays attention to hyper sphere as the space for shapes on more than 2 dimensions.

References

- Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Sciences*, **(1)**, 181-242.
- Bookstein, F. L. (1991). *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, New York.
- Buschang, P. H., De La Cruz, R., Viazis, A. D., and Demirjian, A. (1993). Longitudinal shape changes of the nasal dorsum. *American Journal of Orthodontics and Dentofacial Orthopedics*, **104(6)**, 539-543.
- Clark, R. M. and Thompson, R. (1984). Statistical comparison of palaeomagnetic records from lake sediments. *Geophys. J. R. Astr. Soc.*, **76**, 337-368.
- Di Marzio, M., Panzera, A., and Taylor, C. C. (2013). Nonparametric Regression for Circular Responses. *Scandinavian Journal of Statistics*, **40(2)**, 238-255.
- Fisher, N. I. Lewis, T. and Embleton, B. J. (1987). *Statistical analysis of spherical data*. Cambridge University Press, New York.
- Gould, A. L. (1969). A regression technique for angular variates. *Biometrics*, **25(4)**, 683-700.
- Kendall, D. G. (1977). *The diffusion of shape*. Advances in Applied Probability, **9(3)**, 428-430.
- Kendall, D. G. (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, **16(2)**, 81-121.
- Kent, J. T., Mardia, K. V., Morris, R. J., and Aykroyd, R. G. (2000). Procrustes growth models for shape. In *Proceedings of the First Joint Statistical Meeting, New Delhi, India*, 236-238.
- Le, H., and Kume, A. (2000). Detection of shape changes in biological features. *Journal of Microscopy*, **200(2)**, 140-147.
- Morris, R. J., Kent, J. T., Mardia, K. V., Fidrich, M., Aykroyd, R. G., and Linney, A. (1999). Analysing growth in faces. In *Proc. Conf. Imaging Science, Systems and Technology 1999, Las Vegas*, 404-410.

-
- Parker, R. L., and Denham, C. R. (1979). Interpolation of unit vectors. *Geophysical Journal International*, 58(3), 685-687.
- Prentice, M. J. (1987). Fitting smooth paths to rotation data. *Applied statistics*, 36(3), 325-331.
- Sharma, P., Arora, A., and Valiathan, A. (2014). Age Changes of Jaws and Soft Tissue Profile. *The Scientific World Journal*, 2014.
- Thompson, R., and Clark, R. M. (1981). Fitting polar wander paths. *Physics of the Earth and Planetary Interiors*, 27(1), 1-7.
- Zelditch, M. L., Swiderski, D. L., and Sheets, H. D. (2012). *Geometric Morphometrics for Biologists: A Primer*. Academic Press, New York.



Prediction in GINAR(1) Time Series Model

M. Mohammadpour¹, Kh. Rezvani¹

¹Department of Statistics, University of Mazandaran, Babolsar, Iran

Abstract: The paper focuses on the forecasting of integer-valued time series, modeled by the INAR(1) with geometric innovations. Bayesian methodology is used to obtain point and interval predictions for future values of the process with their classic counterparts. The proposed approaches are illustrated with a simulation study and a real example.

Keywords Bayesian prediction, Count time series, Forecasting, INAR model.

Mathematics Subject Classification (2010): 60G10; 60G25.

1 Introduction

Integer-valued time series can model data with correlated counts are encountered in several life situations, e.g., the number of days with storm, the number of road accidents, the number of foggy days, the number of bases of DNA sequences and so on. The most common integer-valued time series models are constructed via the binomial thinning operator $\alpha \circ$, that was first introduced by Steutel and van Harn (1979) in the form

$$\alpha \circ X = S_X = \sum_{i=1}^X Y_i, \quad \alpha \in (0, 1), \quad (1)$$

where $S_0 = Y_0 = 0$, X is a non-negative integer-valued random variable and $\{Y_i\}$ is a sequence of independent identically distributed (i.i.d) random variables (rvs) with Bernoulli(α) distribution and is independent of X . The first order non-negative integer value autoregressive (INAR(1)) process was first introduced by McKenzie (1985), and Al-Osh and Alzaid (1987) based on the operator " \circ ". Let X_t be a non-negative integer-valued random variable observed at time t , then INAR(1) process is defined as

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t,$$

¹Kh. Rezvani: k.rezvani@stu.umz.ac.ir

where X_t can follow a certain marginal distribution, the innovation (or white noise) $\{\varepsilon_t\}$ is a sequence of i.i.d random variables with some discrete distribution and also $\{\varepsilon_t\}$ is independent of the Bernoulli counting series $\{Y_i\}$ and X_{t-l} , for $l \geq 1$. Modeling of INAR(1) time series was first introduced using Poisson marginal distribution and then considered using geometric marginal distribution (Alzaid and Al-Osh (1988)) and zero truncated Poisson distribution (Bakouch and Ristić (2010)) and et al. For Bayesian part see [5,8].

In this paper, we investigate the forecasting of integer valued AR(1) process with Geometric innovations based on the binomial thinning operator which was modeled by Aghababaei Jazi et al. (2012). Based on the definition of Geometric distribution, INAR(1) process with geometric innovation, denoted by GINAR(1) defined as

$$X_t = \alpha o X_{t-1} + \varepsilon_t, \tag{2}$$

where the binomial thinning operator αo defined above and $\{\varepsilon_t\}$ are iid Ge(p) r.v.'s.

2 Conditional properties of the GINAR(1) model

Let $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$. The h -step ahead conditional expectation and variance of the GINAR(1) process are

$$E(X_{n+h}|\mathbf{X}_n) = \alpha^h X_n + \frac{1 - \alpha^h}{1 - \alpha} \frac{1 - p}{p},$$

and

$$Var(X_{n+h}|\mathbf{X}_n) = \alpha^h (1 - \alpha^h) X_n + \frac{1 - \alpha^h}{1 - \alpha} \frac{1 - p}{p} + \frac{1 - \alpha^{2h}}{1 - \alpha^2} \frac{(1 - p)^2}{p^2},$$

and if $h \rightarrow \infty$, $E(X_{n+h}|\mathbf{X}_n) \rightarrow \frac{1-p}{p(1-\alpha)}$, $Var(X_{n+h}|\mathbf{X}_n) \rightarrow \frac{(p\alpha+1)(1-p)}{p^2(1-\alpha^2)}$, which are the unconditional mean and variance of the process.

In general, the h -step ahead conditional probability function calculates as follows

$$\begin{aligned} f(x_{n+h}|\mathbf{x}_n) &= P(X_{n+h} = x_{n+h} | X_n = x_n) \\ &= P(\alpha^h o X_n + \sum_{i=0}^{h-1} \alpha^i o \varepsilon_{n+h-i} = x_{n+h} | X_n = x_n) \\ &= \sum_{i=0}^{M_h} P(B_{\mathbf{x}_n}^{\alpha^h} = i) P(\sum_{i=0}^{h-1} \alpha^i o \varepsilon_{n+h-i} = x_{n+h} - i), \end{aligned} \tag{3}$$

where notation $B_{\mathbf{x}_n}^{\alpha^h}$ denotes a variable with distribution $Bin(\mathbf{x}_n, \alpha^h)$ and The probability of $W_j = \alpha^j o\varepsilon_{n+h-j}$ obtains as

$$\begin{aligned} P(W_j = w) &= \sum_m P(W_j = w | \varepsilon_{n+h-j} = m) P(\varepsilon_{n+h-j} = m) \\ &= \sum_{m=w}^{\infty} P(\alpha^j o\varepsilon_{n+h-j} = w | \varepsilon_{n+h-j} = m) P(\varepsilon_{n+h-j} = m) \\ &= \sum_{m=w}^{\infty} P(B_m^{\alpha^j} = w) P(\varepsilon_{n+h-j} = m) \\ &= \sum_{m=w}^{\infty} \binom{m}{w} (\alpha^j)^w (1 - \alpha^j)^{m-w} p(1-p)^m. \quad (4) \end{aligned}$$

Regarding the above statements, it can be concluded with regard to the geometry arrivals, $P(\sum_{j=0}^{h-1} W_j)$ can not be accurately obtained. In this case, there is no simplification of the h-step ahead conditional distribution. For $h = 1$, one-step ahead conditional predictive pmf obtained as follows

$$f(x_{n+1} | \mathbf{x}_n) = \sum_{i=0}^{M_1} \binom{x_n}{i} \alpha^i (1 - \alpha)^{x_n - i} p(1-p)^{x_{n+1} - i},$$

where $x_{n+1} = 0, 1, \dots$, and $M_1 = \min(x_{n+1}, x_n)$. For $h = 2$, according to expressions Eqs.(4),

$$\begin{aligned} P(\alpha o\varepsilon_{n+1} + \varepsilon_n = x_{n+2} - i) &= \sum_{l=0}^{x_{n+2}-i} P(\alpha o\varepsilon_{n+1} = l) P(\varepsilon_n = x_{n+2} - i - l) \\ &= \sum_{l=0}^{x_{n+2}-i} \sum_{m=l}^{\infty} P(B_m^{\alpha} = l) P(\varepsilon_{n+1} = m) P(\varepsilon_n = x_{n+2} - i - l) \\ &= \sum_{l=0}^{x_{n+2}-i} p(1-p)^{x_{n+2}-i-l} \sum_{m=l}^{\infty} \binom{m}{l} \alpha^l (1 - \alpha)^{m-l} p(1-p)^m, \end{aligned}$$

and the two-step ahead conditional predictive pmf obtained as follows

$$\begin{aligned} f(x_{n+2} | \mathbf{x}_n) &= \sum_{i=0}^{M_2} P(B_{\mathbf{x}_n}^{\alpha^2} = i) P(\alpha o\varepsilon_{n+1} + \varepsilon_n = x_{n+2} - i) \\ &= \sum_{i=0}^{M_2} \binom{x_n}{i} \alpha^{2i} (1 - \alpha^2)^{x_n - i} \sum_{l=0}^{x_{n+2}-i} p(1-p)^{x_{n+2}-i-l} \\ &\quad \times \sum_{m=l}^{\infty} \binom{m}{l} \alpha^l (1 - \alpha)^{m-l} p(1-p)^m, \end{aligned}$$

where $M_2 = \min(x_{n+2}, x_n)$.

3 Forecasting

In this section, Bayesian methodology is used to obtain point for future values of the process and compare it with their classic counterparts.

3.1 Point Prediction

Classical Methodology

The h-step-ahead predictor based on the conditional expectation of INAR(1) is

$$E(X_{n+h}|X_n) = \alpha^h X_n + \frac{1 - \alpha^h}{1 - \alpha} \frac{1 - p}{p},$$

In order to obtain coherent predictions for X_{n+h} Freeland and McCabe (2003) suggest using the value which minimizes the expected absolute error given the sample, i.e., the value that minimizes $E[|X_{n+h} - \hat{X}_{n+h}| | X_n]$. So $\hat{X}_{n+h} = \tilde{m}_{n+h}$ is the median of the h-step-ahead conditional distribution $f(x_{n+h}|x_n)$.

Bayesian Methodology

Let $\theta \in \Theta$ be the vector of unknown parameters. The h-step- ahead Bayesian posterior predictive distribution is given by

$$f(x_{n+h}|\mathbf{x}_n) = \int f(x_{n+h}|\mathbf{x}_n; \theta) p(\theta|\mathbf{x}_n) d\theta$$

where $p(\theta|\mathbf{x}_n)$ is the posterior probability function of θ and $f(x_{n+h}|\mathbf{x}_n; \theta)$ is the predictive distribution (classical) given in Eqs. (3).

Since beta and gamma are conjugate of binomial and Geometric distributions, respectively, we use them for prior distributions of the parameters of the *INAR*(1) model,

$$p \sim \text{Beta}(a, b), \quad a, b > 0,$$

$$\alpha \sim \text{Beta}(c, d), \quad c, d > 0,$$

The posterior distribution of (α, p) can be written as

$$p(\alpha, p|\mathbf{x}_n) \propto p(\alpha, p) L(\mathbf{x}_n, \alpha, p|x_1)$$

$$\begin{aligned}
&= p^{a-1}(1-p)^{b-1}\alpha^{c-1}(1-\alpha)^{d-1} \\
&\quad \times \prod_{t=2}^n \sum_{i=0}^{M_t} \binom{x_{t-1}}{i} \alpha^i (1-\alpha)^{x_{t-1}-i} p(1-p)^{x_t-i}
\end{aligned}$$

Consequently for the $GINAR(1)$ model, the Bayesian predictive function of x_{n+h} given \mathbf{x}_n is

$$\begin{aligned}
f(x_{n+1}|\mathbf{x}_n) &\propto \int \int \sum_{i=0}^{M_1} \binom{x_n}{i} \alpha^i (1-\alpha)^{x_n-i} \\
&\quad \times p(1-p)^{x_{n+1}-i} p^{a-1}(1-p)^{b-1} \alpha^{c-1} (1-\alpha)^{d-1} \\
&\quad \times \prod_{t=2}^n \sum_{i=0}^{M_t} \binom{x_{t-1}}{i} \alpha^i (1-\alpha)^{x_{t-1}-i} p(1-p)^{x_t-i} d\alpha dp,
\end{aligned}$$

and

$$\begin{aligned}
f(x_{n+2}|\mathbf{x}_n) &\propto \int \int \sum_{i=0}^{M_2} \binom{x_n}{i} \alpha^{2i} (1-\alpha^2)^{x_n-i} \sum_{l=0}^{x_{n+2}-i} p(1-p)^{x_{n+2}-i-l} \\
&\quad \times \sum_{m=l}^{\infty} \binom{m}{l} \alpha^l (1-\alpha)^{m-l} p(1-p)^m p^{a-1} (1-p)^{b-1} \alpha^{c-1} (1-\alpha)^{d-1} \\
&\quad \times \prod_{t=2}^n \sum_{i=0}^{M_t} \binom{x_{t-1}}{i} \alpha^i (1-\alpha)^{x_{t-1}-i} p(1-p)^{x_t-i} d\alpha dp.
\end{aligned}$$

The complexity of $f(x_{n+h}|\mathbf{x}_n)$ does not allow us to work with it directly. In order to estimate X_{n+h} , we can adapt to the integer case of Tanner[10] composition method. To sample $(X_{n+h,1}, X_{n+h,2}, \dots, X_{n+h,m})$, we can use the following algorithm:

Algorithm 1. For sample $(X_{n+h,1}, X_{n+h,2}, \dots, X_{n+h,m})$:

- from the sample (X_1, X_2, \dots, X_n) , calculate (through the classical method) a starting estimative to α , let α_0 ;
- using the adaptive rejection Metropolis sampling (ARMS) within Gibbs methodology [9], calculate from the full conditional distributions of parameters α and λ , a sample $(\alpha_1, \lambda_1), (\alpha_2, \lambda_2), \dots, (\alpha_m, \lambda_m)$;
- For each i ($i = 1, \dots, m$) sample $X_{n+h,i}$ from $f(x_{n+h}|x_n; \alpha_i, \lambda_i)$, using the inverse transform method adapted to integer variables, that is,

- (a) sample u from uniform $U(0; 1)$,
- (b) calculate the least integer valued s : $\sum_{x_{n+h}=0}^s f(x_{n+h}|x_n; \alpha_i, \lambda_i) \geq u$,
- (c) consider $X_{n+h,i} = s$.

After sampling $(X_{n+h,1}, X_{n+h,2}, \dots, X_{n+h,m})$, the h-step-ahead predictor \widehat{X}_{n+h} , can be calculated from sample mean (\widetilde{X}_{n+h}) , median (\widetilde{m}_{n+h}) or mode $(\widetilde{m}o_{n+h})$. But we can also calculate $E(X_{n+h}|X_n)$ as

$$\begin{aligned} E(X_{n+h}|X_n) &= E[E(X_{n+h}|X_n, \theta)|X_n] \\ &= E\left[\alpha^h X_n + \frac{1 - \alpha^h}{1 - \alpha} \frac{1 - p}{p} | X_n\right] \\ &= X_n E[\alpha^h | X_n] + E\left[\frac{1 - \alpha^h}{1 - \alpha} \frac{1 - p}{p} | X_n\right]. \end{aligned}$$

These expected values can be estimated through Markov Chain Monte Carlo (MCMC) algorithms. We perform Metropolis algorithm in conjunction with Adaptive Rejection Sampling Method (ARMS) in order to sample values from full conditional distributions of α and λ ; let them be noted by $(\alpha_1, \alpha_2, \dots, \alpha_m)$, $(\lambda_1, \lambda_2, \dots, \lambda_m)$, respectively. We have

$$\widehat{E}[\alpha^h | X_n] = \frac{1}{m} \sum_{i=1}^m \alpha_i^h, \quad \widehat{E}\left[\frac{1 - \alpha^h}{1 - \alpha} \frac{1 - p}{p} | X_n\right] = \frac{1}{m} \sum_{i=1}^m \frac{1 - \alpha_i^h}{1 - \alpha_i} \frac{1 - p_i}{p_i},$$

consequently the predictor can be written as

$$\widehat{X}_{n+h} = X_n \left(\frac{1}{m} \sum_{i=1}^m \alpha_i^h \right) + \left(\frac{1}{m} \sum_{i=1}^m \frac{1 - \alpha_i^h}{1 - \alpha_i} \frac{1 - p_i}{p_i} \right).$$

3.2 Interval Prediction

The procedure used in this section is an adaptive generalization of the method used to obtain HPD (Highest Posterior Density) intervals of the model parameters, in which we consider the predictive distribution instead of the posterior.

Definition 1.2 $R(\gamma) = (X_L, X_R)$ is a HPD interval for X_{n+h} if

$$P(X_L < X_{n+h} < X_R) = \sum_{x_{n+h}=X_L}^{X_R} f(x_{n+h}|\mathbf{x}_n) \geq k_\gamma, \tag{5}$$

where k_γ is the largest constant such that $P(X_{n+h} \in R(\gamma)) \geq \gamma$.

Due to complexity of the predictive probability function given it is not possible to calculate the exact HPD interval for X_{n+h} ; we can give an approximation for $R(\gamma)$ by using the Chen and Shao [6] algorithm, because this method does not require the knowledge of the closed form of $f(x_{n+h}|x_n)$. The Chen and Shao algorithm, can be described as,

Algorithm 2. *The Chen and Shao algorithm has the following steps:*

- Obtain an MCMC sample $(X_{n+h,1}, X_{n+h,2}, \dots, X_{n+h,m})$ (Algorithm 1.);
- consider $(X_{(n+h,1)}, X_{(n+h,2)}, \dots, X_{(n+h,m)})$;
- compute the credible intervals $R_i(\gamma) = (X_{(n+h,i)}, X_{(n+h,i+[m\gamma])})$, $1 < i < m - [m\gamma]$; where $[m\gamma]$ is integer part of $m\gamma$;
- the HPD interval to X_{n+h} is the one, denoted by $\widehat{R}(\gamma)$, with the smallest amplitude among all credible intervals.

Sometimes we obtain more than one interval. For this situation, we consider for $\widehat{R}(\gamma)$ the interval with greater absolute frequency, among the smaller intervals width. When the interval is still not unique we take the one with the smallest lower limit of the interval.

3.3 Some simulation results

For the simulation study we consider samples with size $n = 50, 100, 200$ generated by GINAR(1) models with the parameters values $\alpha = 0.2, 0.8$ and $p = 0.3, 0.6$.

From the various simulated samples we conclude that large values α and small values p are related with high dispersion values. Consequently the increase in α and decrease in p provides large values of $|X_{n+h} - X_n|$. To confront classical and Bayesian methodologies we use the mean square error (MSE) to compare means, the mean absolute deviation (MAD) to compare medians and the "everything or nothing" lost function (FPTN)[8], given by $\frac{1}{n} \sum I(x_{n+h})$ where

$$I(x_{n+h}) = \begin{cases} 1, & |\widehat{x}_{190+h} - x_{190+h}| > \delta \\ 0, & |\widehat{x}_{190+h} - x_{190+h}| \leq \delta \end{cases}$$

to compare modes. In this situation we consider $\delta=1$ since we have integer values.

Table 1 shows the MSE, MAD and FPTN values from 10 predictions h-step-ahead for three samples with sizes 40, 90 and 190 of this model. This table is presented considering $p=0.3$, but for $p=0.6$ the conclusions are similar. Values of MSE, MAD and FPTN were calculated, respectively, through means, medians and modes obtained by Algorithm 1. The indices "C" or "B" indicate which methodology is used (classical or Bayesian, respectively). Table 2 shows the coverage probability estimates and mean amplitudes of the intervals for the future values, in GIAR(1) model with 100 repeat and $p=0.3$.

4 Real data analysis

In this section, we discuss some possible applications of the GIAR(1) model for real count time series data.

The data give numbers of submissions to animal health laboratories, monthly 2003–2009, from a region in New Zealand. The submissions contains many categories for presenting symptoms. After checking autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) of the series, the first order autoregression models are appropriate for the data series. We fitted GIAR(1) model to the data by conditional maximum likelihood method as

$$X_t = 0.3194476 \circ X_{t-1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ are iid $\text{Ge}(0.6161064)$ r.v.'s. The classic and bayesian prediction values h-step-ahead ($h=1, \dots, 10$) prediction for real data time series are provided in Tables 3. Their MSE and FPTN values for classic are 0.7228 and 1.8790 and for bayesian are 0.7616 and 1.9920.

Table 1: MSE, MAD and FPTN values of h-step-ahead (h=1), considering samples with sizes 40, 90 and 190 simulated from GINAR(1) model with 100 repeat,(p = 0.3).

α	0.2			0.8		
n	40	90	190	40	90	190
MSE_C	8.75	7.32	7.84	14.26	15.80	14.48
MSE_B	6.66	7.37	7.81	14.01	15.82	14.38
MAD_C	2.26	1.97	0.112	0.269	0.246	0.281
MAD_B	0.115	0.130	0.130	0.269	0.254	0.281
$FPTN_C$	0.121	0.114	0.179	0.289	0.275	0.300
$FPTN_B$	0.118	0.199	0.169	0.306	0.292	0.319

Table 2: Amplitude and Coverage probability with $p = 0.3$.

α	0.2				0.8			
	cov. prob		mean. amp		cov. prob		mean. amp	
h	C	B	C	B	C	B	C	B
1	0.79	0.77	9.40	9.20	0.95	0.91	11.84	10.74
2	0.81	0.79	9.85	9.84	0.87	0.75	16.67	14.94
3	0.79	0.76	9.74	9.74	0.87	0.74	16.60	13.86
4	0.79	0.76	9.72	9.37	0.82	0.74	16.60	14.22
5	0.79	0.79	9.71	9.20	0.82	0.69	16.86	14.14
6	0.81	0.82	9.71	9.24	0.77	0.71	16.61	14.20
7	0.76	0.73	9.67	9.09	0.73	0.65	16.06	14.98
8	0.77	0.74	9.67	9.96	0.73	0.64	16.05	12.36
9	0.79	0.78	9.63	9.37	0.73	0.68	16.77	13.41
10	0.66	0.65	9.65	9.23	0.83	0.72	16.16	12.98

Table 3: Prediction values of h-step-ahead (h=1,...,10) prediction for real data time series

h	1	2	3	4	5	6	7	8	9	10
$mean_C$	0.6247	0.8178	0.8775	0.8959	0.9016	0.9033	0.9040	0.9041	0.9041	0.9041
$mean_B$	0.6961	0.8644	0.9103	0.9240	0.9284	0.9299	0.9305	0.93068	0.9307	0.9308

References

- Aghababaei Jazi, M., Jones, G., & Lai, C. D. (2012). Integer valued AR (1) with geometric innovations. *Journal of Iranian Statistical Society*, 11(2), 173-190.
- Al-Osh , M. A. and Alzaid, A. A. (1987). Firs-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, 8 (3), 261-275.
- Alzaid, A., & Al-Osh, M. (1988). First-Order Integer-Valued Autoregressive (INAR (1)) Process: Distributional and Regression Properties. *Statistica Neerlandica*, 42(1), 53-61.
- Bakouch, H. S., & Ristic, M. M. (2010). Zero truncated Poisson integer-valued AR (1) model. *Metrika*, 72(2), 265-280.
- Freeland, R. K., & McCabe, B. P. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20(3), 427-434.
- Gilks, W. R. , Best, N. G. and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 455-472.
- McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, 21 (4), 645-650 .
- Silva, N., Pereira, I. and Silva, M. E. (2009). Forecasting in INAR (1) model. *REVSTAT Statistical Journal*, 7(1), 119-134.
- Steutel, F. W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 893-899.
- Tanner, M. A. (1991). *Tools for statistical inference*, 3. New York: Springer .



Notes on Generalized Reversed Failure Rate

G. R. Mohtashami Borzadaran¹

¹Department of Statistics Faculty of Mathematical Sciences Ferdowsi University of
Mashhad

Abstract: There has been growing interest in study of reliability function such as reversed failure rate and its properties and applications. Some of them are similar to those of failure rate and some have different approach and motivation. Distribution with an increasing (decreasing) generalized failure rate IGFR (DGFR) has useful applications in pricing and supply chain problems. Many distributions with their GFR and the closure properties of them have important role in economical aspects. Some notion of ageing is related to generalized reversed failure rate with the implications and view to the link with elasticity function is obtained and mentioned shortly. The DGRFR and related version are studied via a variant point of view criteria and also discussed closure properties. The weighted distributions with monotone generalized reversed failure rate is our concentration for finding more approach and result based on the GRFR of original distribution which is the last part of the paper.

Keywords Elasticity function, Reversed failure rate, Increasing generalized failure rate, Generalized reversed failure rate, Weighted distributions.

Mathematics Subject Classification (2010): 60E05 60E10 62N99.

1 Introduction

It is well-known fact that besides the distribution of random variable can be defined using the cumulative distribution function, the probability density (mass) function or characteristic function. Other functions such as survival function, failure rate, reversed failure rate and mean residual life functions are those applicable in reliability. An important phenomenon in reliability theory is ageing. Measures of ageing are like the above functions with their major role in life sciences. The concept of failure rate

¹G. R. Mohtashami Borzadaran : gmb1334@yahoo.com

is very well-known in the reliability literature. We know that the failure rate (FR) function can be interpreted as the probability of failure in an infinitesimal unit interval of time. This interpretation and some other properties such as increasing failure rate (IFR) of an object is an ageing of some kind, which is an important property in various applications. Many engineering items are described by an increasing failure rate. The reversed failure rate (RFR) defined as the ratio of the density to distribution function which has attracted the attention of the researchers. [Keilson and Sumita \(1982\)](#) were among the first to define the RFR and call it dual failure function. It is known ([Block et al. \(1998\)](#)) that in the case of absolutely continuous distributions on the positive real line there does not exist a model with constant reversed hazard rate. There has been growing interest in recent times in the study of reliability functions in reversed time and their applications. Some analogous results and characterizations in the case of the reversed hazard rate and mean residual life for discrete distributions are discussed in recent years. The study of the monotonicity of the RFR and FR functions, known in the literature as ageing notions with applications in different fields, such as economic analysis, actuarial science, reliability theory, among others. Later authors such as [Chechile \(2011\)](#) studied more on reversed failure rate and recently [Torrado and Oliviera \(2013\)](#) obtained results on proportional reversed failure rate class. Analogously, introduced the generalized reversed failure rate (GRFR) (called elasticity) as the variant extension of the RFR. It should be noted that the elasticity of the distribution of a continuous and non-negative random variable defined in [Veres-Ferrer and Pavia \(2014\)](#) which is the GRFR. In addition, the decreasing generalized reversed failure rate (DGRFR) class has been recently used in order to compare stochastically order statistics in the scale model [see [Khaledi et al. \(2011\)](#)]. Moreover, they showed that generalized gamma and power-generalized Weibull distributions with arbitrary parameters belong to the DGRFR class.

In this paper, review on reversed hazard rate and its extensions, some properties of GRFR, IGRFR and DGRFR are obtained. Also their connection with elasticity function and application in economics is investigated. The link and notes in view of weighted distributions is the last part of the paper. Note that the discrete case is the future of the work.

2 Decreasing Reversed Failure Rate

Suppose the distribution of a continuous random variable X has the probability density function f (pdf) and cumulative density function F (cdf), the survival function is given by $S(x) = P(X > x) = 1 - F(x) = \bar{F}(x)$ which is monotonically decreasing. The hazard (failure) rate function and the reverse hazard (failure) rate function of X are given by $r(x) = \frac{f(x)}{F(x)}$ and $\tilde{r}(x) = \frac{f(x)}{\bar{F}(x)}$ respectively. If $r(x)$ is monotonically increasing (decreasing), then the distribution of X has an increasing (decreasing) failure rate, denoted by IFR (DFR). If $r(x)$ is constant, then the distribution of X has a constant failure rate. The reversed (reverse) hazard (failure) rate, also named retro hazard, was first mentioned by the name dual of the hazard (failure) rate'. The name reversed hazard (failure) rate was first used by [Lagakos et al. \(1988\)](#). The reversed failure rate function has many properties that are similar to those of the usual failure rate function. The decreasing reversed failure rate (DRFR) and increasing reversed failure rate (IRFR) have important role in engineering and economics problems. The following definitions for the DRFR and IRFR classes is our need in this note:

Definition 2.1. *A non-negative random variable X , having distribution F , is said to have DRFR (IRFR), if the function $r^\sim(t), t > 0$ is decreasing (increasing).*

A wide range of distributions belong to the DRFR class. For example, the exponential distribution is a DRFR; Weibull and gamma with shape parameters less than 1 are also DRFR. Not only this, but also increasing failure rate IFR distributions like Weibull, gamma and log normal distributions were found to be DRFR distributions. For details, we refer to [Block et al. \(1998\)](#). It has been proved mathematically that a decreasing failure rate (DFR) distribution is necessarily DRFR distribution. [Block et al. \(1998\)](#) have noted the interesting fact that there is no life distribution which belongs to the class IRFR over the domain $[0, \infty)$. The following assertions are noticeable :

- F has a decreasing reversed failure rate $\iff F$ is log-concave.
- Let $r(t) = c$ (constant), then $\tilde{r}(t) = \frac{c}{e^{ct}-1}$ which is decreases as $t \rightarrow \infty$.
- $r(t)$ decreasing, then $\tilde{r}(t)$ is also decreasing. Also, the increasing RFR on $[a, \infty)$, $\forall a \geq 0$ does not exist. It tends to the pdf $f(x)$ as $x \rightarrow \infty$.

- $X \sim \beta \frac{t^{\beta-1}}{\theta^\beta} \exp\{-(\frac{t}{\theta})^\beta\}$, $t > 0$ (Weibull distribution) has DRFR property.
- It is well known that mixtures of decreasing failure rate (DFR) distributions are always decreasing. It turns out that very often mixtures of increasing failure rate (IFR) distributions can decrease as well.
- If X is DFR (DRFR) then X^b is DFR (DRFR) for $b \geq 1$. It can be seen that X^b is IFR if $0 < b \leq 1$. Also DFR and DRFR distributions are closed under shifting.

3 Generalized Reversed Failure Rate

As we know, introduced the generalized failure rate in Lariviere and Porteus (2001) and length-biased failure rate in Righter et al. (2009). Veres-Ferrer and Pavia (2014) motivated by the recent use of the elasticity and GRFR in economics and in reliability stochastic comparisons among systems by Khaledi et al. (2011). Analogous, the definition of generalized reversed failure rate (GRFR) of the random variable X with pdf f and cdf F as $\tilde{h}(x) = x\tilde{r}(x) = \frac{xf(x)}{F(x)}$ leads to an interesting idea. Some of the results and notes due to GRFR with one application in economics as elasticity function is mentioned below:

- Let $X \sim GG(p, q)$, $p, q > 0$, then $\tilde{h}(x) (x^2 r'(x))$ is decreasing (increasing) function of x .
- For any non-negative random variable with domain $[c, b]$, $0 < c < b < \infty$ and mean μ and variance σ^2 : $E(\tilde{h}(X)) > \frac{2\mu^2}{b^2 - (\mu^2 + \sigma^2)}$, equality does not hold, there is not a constant B such that $\sqrt{x F(x)} = B \sqrt{x \frac{f^2(x)}{F(x)}}$. Similar inequality hold for GFR and equality holds for exponential distribution. Also note that $E(h(X)) = \alpha E(Xr(\alpha X))$ seems does not correct.
- For any continuous probability distribution, the cumulative probability distribution and probability density are $F(x) = \frac{h(x)}{h(x)+\tilde{h}(x)}$ and $f(x) = \frac{h(x)\tilde{h}(x)}{x(h(x)+\tilde{h}(x))}$ respectively.
- For continuous probability distribution with parameter θ , Fisher information is $I_X(\theta) = E[\frac{\partial \ln f(X, \theta)}{\partial \theta}]^2$ but note that we can have this statement based on GRFR as $I_X(\theta) = E[\frac{\partial \ln \tilde{h}(X, \theta)}{\partial \theta}]^2$ and the Shannon entropy is achieved via $h(X) = E(-\ln f(X, \theta)) = E(\ln X) - E(\ln \tilde{h}(X, \theta)) + 1$

- The elasticity of a function of a single variable where f is a differentiable real-valued function of a single variable. Define the elasticity of $f(x)$ with respect to x (at the point x) to be $E(x) = \frac{xf'(x)}{f(x)}$ or another version $E(x) = \frac{\frac{df(x)}{dx}}{\frac{f(x)}{x}}$ which show that the elasticity of $f(x)$ with respect to x is the ratio of the percent change in $f(x)$ to the corresponding percent change in x . The elasticity of the function f at a point of x can also be thought of as the slope of a graph that plots $\ln x$ on the horizontal axis and $\ln f(x)$ on the vertical axis that is $E(x) = \frac{d \ln f(x)}{d \ln x}$. It is interesting to consider the special case where the elasticity of $f(x)$ with respect to x is a constant. In this case, we have $\ln f(x) = e \ln x + a$ where e and a are constants. Thus we see that f has constant elasticity if and only if f is a power function. In general, the elasticity of f with respect to x depends on the value of x . For example if $f(x) = a - bx$, then $E(x) = \frac{-bx}{a-bx}$. Suppose that the function f is either strictly increasing or strictly decreasing. Then there is a well defined inverse function ϕ , defined so that such that $\phi(y) = x$ if and only if $f(x) = y$. It turns out that if $E(x)$ is the elasticity of $f(x)$ with respect to x , then $\frac{1}{E(x)}$ is the elasticity of $\phi(y)$ with respect to y .

- One of the notion of elasticity function has application in economics. For demand function $Q = f(P)$, the elasticity is defined as $\frac{\frac{\Delta Q}{Q}}{\frac{\Delta P}{P}}$. If the change in P is infinitesimal, the expressions ΔP and ΔQ reduces to differential of dP and dQ respectively which leads to point elasticity of demand. It is valid not only for demand function but also for other functions. The most common applications of the aspects of elasticity of demand is to monopoly theory, where a monopolist is selling a good and the quantity of the good that is demanded is a function $D(p)$ of the monopolists price p . The monopolists revenue is $R(p) = pD(p)$. Does the monopolists revenue increase if increases his price and how is this related to the price elasticity? We note that $R(p)$ is increasing (decreasing) in price if and only if $\ln R(p)$ increases (decreases) as the log of price increases. But $\ln R(p) = \ln p + \ln D(p)$. Then $\frac{d \ln R(p)}{d \ln p} = 1 + E(p)$ where $E(p)$ is the price elasticity of demand. So revenue is an increasing function of p if $E(p) > 1$ (demand is inelastic) and a decreasing function of p if $E(p) < 1$ (demand is elastic).

4 Decreasing Generalized Reversed Failure Rate

Lariviere and Porteus (2001) introduced GFR for continuous random variable. Determining with increasing property of GFR as mentioned in Paul (2005), Lariviere (2006) with its application in economics problems. Also, GRFR defined via similar arguments as in Torrado and Oliviera (2013). It has connection with elasticity function and so with economics. The IGRFR (DGRFR) has important role in risk and economics aspects such as demand. The various properties of IGRFR (DGRFR) and some of the link with other related criterion are as follows:

- Note that, as Marshall and Olkin (2007) proved (see p. 179), non-negative random variables cannot have distributions with increasing reversed failure rate (IRFR). But for distributions with support of the form (∞, a) , with $a < \infty$, one can construct an IRFR distribution, see Block et al. (1998). Therefore, it follows that one can obtain an IGRFR distribution with bounded support of the form $(0, b)$, with $b = e^a < \infty$.
- DGRFR distributions are closed under convolution. Also, this new ageing notion is related to IFR (DFR) and IGFR (DGFR). Also, if X is DGFR (IGFR) then it is also DGRFR (IGFR).
- The below statements are equivalent:
 X is *DGRFR* $\iff \log X$ is *DRFR* $\iff X \leq_{rh} \lambda X$ for $\lambda \geq 1$ $\iff \log X \leq_{rh} a + \log X$ for $a \geq 0$ $\iff \lambda X$ is *DGRFR* for $\lambda \geq 0$. Note that $X \leq_{rh} Y$ if the RFR of X is greater than or equal to the RFR of Y for each element of the domain of these two random variables.
- IGRFR distributions are also closed under positive scale transformations. Moreover, it is easy to see that X is IGRFR if and only if $\log X$ is IRFR.
- Let $X \sim \text{Beta}(\alpha, \beta)$, $\alpha, \beta > 0$. For any $\alpha > 0$, if $\beta \leq 1$ then X is IGRFR, while if $\beta \geq 1$, X is DGRFR.
- X has an increasing (decreasing) generalized failure rate, denoted by IGFR (DGFR), (or equivalently, F is an IGFR (DGFR) distribution) if $h(x)$ is increasing (decreasing).

- X has an increasing (decreasing) generalized reversed failure rate, denoted by IGRFR (DGRFR), (or equivalently, F is an IGRFR (DGRFR) distribution) if $\tilde{h}(x)$ is increasing (decreasing).
- Let X be a non-negative random variable with distribution function F . Then, X is IGRFR (DGRFR) if and only if $F(e^x)$ is log-convex (log-concave).
- If $X_{(t)}$ (X) is DGFR (DRFR) then X ($X_{(t)}$) is IGRFR (DRFR, also IFR).
- Let X be a non-negative random variable with distribution function F . Then, X is IGRFR (DGRFR) if and only if $F(e^x)$ is log-convex (log-concave).
- Let X be a continuous DGRFR random variable with support on the interval $[L, U] \subseteq \mathfrak{R}$. Then for truncated distribution of X which is defined on $[a, b] \subseteq [L, U]$ is also DGRFR. Also, for any $a \in (L, U]$, then the GRFR of the random variable X and its truncation distribution on interval $[L, a]$ are the same.
- Suppose that $\psi(x) = \frac{\phi'(x)}{\phi(x)}$ is a decreasing function.
 - (i) If X is DGRFR (DGFR) then $Y = \phi^{-1}(x)$ is also DGRFR (DGFR).
 - (ii) If X is IGRFR (IGFR) then $Y = \phi^{-1}(x)$ is also IGRFR (IGFR).
 Suppose that ϕ is a concave function.
 - (i) If X is DFR (DRFR) then Y is also DFR (DRFR) and ϕ is a convex function implies
 - (ii) if X is IFR (IRFR) then Y is also IFR (IRFR).
- X is DGRFR (DGFR) then X^b , ($b > 0$) and $X + a$, ($a > 0$) are DGRFR (DGFR).
- The DGRFR distributions are closed under convolutions.
- For continuous probability distribution with $f(U) = 0$ and $F(U) = 1$, the GRFR is monotonically decreasing if GFR is monotonically increases. Also, the GRFR is monotonically decreasing if the density function is monotonically decreasing.
- Give $\eta(x) = x \frac{f'(x)}{f(x)}$, then $\tilde{h}(x)$ is increasing (decreasing) when $\tilde{h}(x) < (>) 1 + \eta(x)$.

5 Generalized Reversed Failure Rate for Weighted Distributions

When an investigator records an observation by nature according to a certain stochastic model, the recorded observation will not have the original distribution unless every observation is given an equal chance of being recorded. We examine a general model leading to weighted distributions with weight functions. Let X be a random variable with pdf f and suppose $X = x$ occurs, the probability of recording it is $w(x)$ depending on the observed value x and unknown parameter α . The pdf of the random variable X^w is $f_w(x) = \frac{w(x,\alpha)f(x,\theta)}{E(w(X,\alpha))}$, where $0 < E(w(X,\alpha))$ exists (see Rao (1965)). We are concentrated on discussing failure rate and reversed failure rate which are in Bartoszewicz and Skolimowska (2006) and variant version of them such as increasing (decreasing) behavior of GFR and GRFR for weighted distributions. Consider the generalized reversed failure rate (GRFR) as $\tilde{h}(x) = x\tilde{r}(x)$ and the GRFR of weighted distribution increasing weight $w(X) = U$ via the statements, $\tilde{h}_w(x) = x\frac{f_w(x)}{F_w(x)} = \frac{w(x)}{E(U)}\tilde{h}(x)\frac{F(x)}{L_U(F(x))}$, where $\frac{F(x)}{L_U(F(x))}$ is decreasing and $L_U(F(x)) = \frac{1}{E(U)}\int_0^{F(x)} wF^{-1}(z)dz$ (Lorenz curve). For decreasing weight we have the above phrase with $\frac{F(x)}{L_U(F(x))}$ (which is increasing) in place of $\frac{F(x)}{L_U(F(x))}$.

- We can find weighted version of GRFR based on GFR on using decreasing weight and $\tilde{h}_w(x) = \frac{w(x)}{E(U)}h(x)\frac{\bar{F}(x)}{L_U(F(x))}$. For increasing weight we have the above phrase with $\frac{\bar{F}(x)}{L_U(F(x))}$ (which is decreasing) in place of $\frac{F(x)}{L_U(F(x))}$.
- Let w be a monotone left continuous function. If w is increasing (decreasing) and $w(x)\tilde{h}(x)$ is decreasing (increasing), then F_w is GRFR (IRFR and IGFR). If $w(x)h(x)$ is decreasing (increasing), then F_w is DGRFR.
- We can apply for the GRFR property in some special cases of the weighted distributions such as the weight $w(x) = e^{lx}x^i[F(x)]^j[\bar{F}(x)]^k$ is one that implies many famous weights. If $(i = j = k = 0)$, $(l = j = k = 0)$, $(i = l = 0, k = n - j)$, $(i = l = k = 0)$ and $(i = j = l = 0)$, then $w(\cdot)$ is moment generating function, moments, order statistics, reversed proportional hazard and proportional hazard respectively. Also, weight $w(x) = [-\ln F(x)]^j[-\ln \bar{F}(x)]^k$ is another weight that $(j = 0)$ and $(k = 0)$ implies upper record and lower record respectively.

References

- Bartoszewicz, J., Skolimowska, M. (2006). Preservation of classes of life distributions and stochastic orders under weighting. *Statist. Probab. Lett.* **76**, 587-596.
- Block, H. W., Savit, T. H., Singh, H. (1998). The reversed hazard rate functions, *Probability in the Engineering and Informational Sciences*, **12**, 69-90.
- Chechile, R. A. (2011). Properties of reversed hazard functions, *Journal of Mathematical Psychology*, **55**, 203-222.
- Keilson, J., Sumita, U. (1982). Uniform stochastic ordering and related inequalities, *Canadian Journal of Statistics*, **10**, 181-198.
- Khaledi, B.-E., Farsinezhad, S. Kochar, S. C. (2011). Stochastic comparisons of order statistics in the scale model, *Journal of Statistical planning and Inference*, **141**, 276286.
- Lagakos, S. W., Barraj, S. M., De Grutolla, V. (1988). Nonparametric analysis of truncated survival data with application to AIDS, *Biometrika*, **75**, 515-523.
- Lariviere, M. A., Porteus, E. L. (2001). On Selling to newsvendor : an analysis of price only contracts, *Manufacturing Service Oper. Management* , **3(4)**, 293-305.
- Lariviere, M. A. (2006). A note on probability distributions with increasing generalized failure rates, *Operations Reseach*, **Vol. 54, No. 3**, 602-604.
- Marshall, A. W., Olkin, I. (2007). Life Distributions: Structure of Nonparametric, Semiparametric and Parametric Families, *Springer, New York*.
- Paul, A. (2005). A note on closure properties of failure rate distributions, *Operations Reseach*, **Vol. 53, No. 4**, 733-734.
- Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment, in *Classical and Contagious Discrete Distributions*, G. Patil Ed., Pergamon Press and Statistical Publishing Society, Calcutta, 320-332.

-
- Righter, R., Shaked, M., Shantikumar, J. G. (2009). Intrinsic ageing and classes of nonparametric distributions, *Probability in the Engineering and Informational Sciences*, **23(4)**, 563-582.
- Torrado, N., Oliveira, P. E. (2013). On closure properties of risk aversion measures, *Preprint No. 13-45*, *Universidade de Coimbra*, 1-21.
- Veres-Ferrer, E. J., Pavia, J. M. (2014). On the relationship between the reversed hazard rate and elasticity, *Statistical Papers*, Doi 10.1007/s00362-012-047-1.



Randomized order restricted estimators for treatment means with a control group

R. Momeni¹, J. Etminan²

¹Department of Statistics, University of Birjand

²Department of Statistics, University of Birjand

Abstract: We consider an experiment which consists of k treatment groups and a control group. Let the sample means $\bar{X}_0, \dots, \bar{X}_k$ be independent normal variates with expected values μ_0, \dots, μ_k and with variances $\frac{\sigma^2}{w_0}, \dots, \frac{\sigma^2}{w_k}$. Let w_0, \dots, w_k be positive weights and μ_0^*, \dots, μ_k^* be the restricted maximum likelihood estimators (RMLE) subject to the constraints $\mu_0 \leq \mu_i$ for $1 \leq i \leq k$. Lee (1988) establish that for large k , the RMLE fails in the sense of mean squared error. Since we propose the new estimator that uniformly performs better than unrestricted and restricted maximum likelihood estimators. Our simulation study suggests that the our estimator compete well with two another estimators. Although the gains in efficiency due to the procedures proposed can be substantial.

Keywords: Randomized order restricted estimator (RORE), Restricted maximum likelihood estimator (RMLE), Unrestricted maximum likelihood estimator (UMLE), Mean squared error (MSE).

1 Introduction

A typical problem of interest is to compare the means of the treatment groups with that of the control group. Let the sample means $\bar{X}_0, \dots, \bar{X}_k$ be independent normal variates with unknown means μ_0, \dots, μ_k satisfying $\mu_0 \leq \mu_i$ for all $i \geq 1$ and $var(\bar{X}_i) = \sigma_i^2 = \frac{\sigma^2}{w_i}$, whenever w_i be positive weights. This property is known as the simple tree ordering (Barlow, et al. (1972)). In this article, we shall assume that the simple tree ordering is our prior knowledge. This priori information arises very naturally in many problems of practical interest. For example, suppose μ_0 is the average yield of a crop with no fertilizer added and $\mu_i, i \geq 1$, is the average yield of the crop when the i th brand of fertilizer

¹R. Momeni: Rm1360@gmail.com

is added. For an excellent review on this subject one may refer to the recent book by Silvapulle and Sen (2005). It is well known that the restricted maximum likelihood estimator (RMLE) of μ_0, \dots, μ_k under some inequality restrictions, may perform poorly in terms of mean squared error (Lee (1988)). Let $\bar{X}_0, \dots, \bar{X}_k$ be the sample means for the $k + 1$ groups, since they are the unrestricted maximum likelihood estimator (UMLE) for μ_0, \dots, μ_k . The UMLEs are unbiased but they may fail to satisfy simple tree ordering, $\mu_0 \leq \mu_i$. Lee (1988) introduced the estimator with smaller mean squared error point wise than the usual estimator $(\bar{X}_0, \dots, \bar{X}_k)$, whereas this methodology utilized the prior knowledge $\mu_0 \leq \mu_i; i \geq 1$. This procedure as follows: let w_0, \dots, w_k be positive weights and $\mu^* = (\mu_0^*, \dots, \mu_k^*)$ be the weighted least squares estimator, i.e., μ^* minimizes $\sum_{i=0}^k (\bar{X}_i - \mu_i)^2 w_i$ subject to the restriction simple tree ordering $\mu_0 \leq \mu_i$. The weighted least squares estimator μ^* is known as the isotonic regression under the simple tree ordering (Barlow, et al. (1972)). For the special case when the natural weights $w_i = n_i$ are used, the isotonic regression μ^* is the maximum likelihood estimator subject to the simple tree order restriction. Since the isotonic regression estimator can be expressed in a nice form using the minimax formula given on Robertson, Wright and Dykstra (1988). In this article, as a first step we focus on the isotonic regression estimator, typically the restricted maximum likelihood estimator (RMLE) and in Section 2, we discuss the problem of estimating the parameters under tree order restriction. In Section 3, we discuss the problem of estimating the control group mean parameter under tree order restriction. Surprisingly, however, the proposed estimator dominates the unrestricted maximum likelihood estimator (UMLE). Simulation results and performance of the proposed estimation procedure is numerically evaluated in Section 4. Conclusion of this article presented in Section 5 with a brief summary and a discussion of some interesting open research problems in this area.

2 Restricted maximum likelihood estimator (RMLE)

To describe the isotonic regression estimator, we let " \leq " denote the pair relation between i and j , both in $\tau = \{0, 1, \dots, k\}$, such that $i \leq j$ if and only if it is known that $\mu_i \leq \mu_j$. Then the minimax formula for estimating μ_i is:

$$\hat{\mu}_i = \min_{L:i \in L} \max_{U:i \in U} A_{\bar{X}}(L \cap U) = \max_{U:i \in U} \min_{L:i \in L} A_{\bar{X}}(L \cap U) \quad (2.1)$$

Where U and L are respectively, the upper and lower sets of τ . An upper set U is a set such that if $i \in U$ and $i \leq j$ then $j \in U$. A lower set L is defined similarly with $i \leq j$ replaced by $j \leq i$. Furthermore, for an arbitrary set S ,

$$A_{\bar{X}}(S) = \frac{\sum_{i \in S} w_i \bar{X}_i}{\sum_{i \in S} w_i} \quad (2.2)$$

For simple tree order restricted, expression (2.1) reduces to:

$$\hat{\mu}_0^{ST} = \min_{0 \in S} A_{\bar{X}}(S) \quad (2.3)$$

Where S is any subset of $\{0, 1, \dots, k\}$ containing the element 0, where the minimization is taken over all S sets, since:

$$\hat{\mu}_i^{ST} = \max(\hat{\mu}_0^{ST}, \bar{X}_i); \quad \text{for all } i \geq 1 \quad (2.4)$$

Lee (1988) established that for large k ,

$$E(\mu_0^* - \mu_0)^2 > E(\bar{X}_0 - \mu_0)^2 \quad (2.5)$$

Therefore the reverse inequality (2.5) under the simple tree ordering is the counterexample in the literature. Since knowing that simple tree ordering ($\mu_0 \leq \mu_i; i \geq 1$) holds does not guarantee the mean squared error reduction for the control group. On the other hand, theorem 2.1 (Lee (1988)) provides a counterexample by illustrating the unboundedness of the bias of μ_0^* .

3 The proposed methodology

3.1 Motivation

Let X_{ij} , $i = 0, 1, \dots, k$ and $j = 1, 2, \dots, n_i$ denote the j th observation from the i th dose group, where $X_{ij} \stackrel{iid}{\sim} N(\mu_i, \sigma_i^2)$ with $\mu_0 \leq \mu_i$, for all $i \geq 1$. Let \bar{X}_i denote the sample means for the i th group, where $i = 0$ refers to the control group, then \bar{X}_i is the unrestricted maximum likelihood estimator (UMLE) of μ_i . On the other hand, under the simple tree order restriction $\mu_0 \leq \mu_i$, $i \geq 1$, then restricted maximum likelihood estimators (RMLE) of μ_0, \dots, μ_k are given by:

$$\tilde{\mu}_0 = \min_S \left(\frac{\sum_{i \in S} w_i \bar{X}_i}{\sum_{i \in S} w_i} \right) \quad (3.1)$$

$$\tilde{\mu}_i = \max(\bar{X}_i, \tilde{\mu}_0); \quad i = 1, 2, \dots, k \tag{3.2}$$

Where S denotes a subset of $\tau = \{0, 1, \dots, k\}$ which contains 0. So, we can interpret $\tilde{\mu}_0$ as follows. Maintaining all the known inequalities $\mu_0 \leq \mu_i, i \geq 1$, construct all $k!$ simple order restrictions between μ_0, \dots, μ_k by considering all $k!$ ordering between μ_1, \dots, μ_k . Under each such simple order restriction obtain the corresponding RMLE. Then $\tilde{\mu}_0$ is the minimum among all such RMLEs. Lee (1988) observed and established that, as k increases, the mean squared error of $\tilde{\mu}_0$ would grows larger than that of the UMLE. Thus, it would fail to dominate the UMLE in terms of mean squared error.

3.2 Randomized order restricted estimator (RORE)

In this Section, we propose the following estimator for constraint parameter space, that is $\mu_0 \leq \mu_i$ for all $i \geq 1$. To motivate the development of some alternate estimators to the restricted maximum likelihood estimators, we first consider the estimation of the smallest parameter, μ_0 , in a simple tree ordering. As regards, probability of appropriation depends on the number of treatments and it's a nondecreasing function of k , Therefore according to considerations in Section 2, we proposed the randomized order restricted estimator, that is:

$$\hat{\mu}_0 = \begin{cases} \bar{X}_0 = \text{UMLE}; & \text{with probability } p = \frac{k}{k+1} \\ \tilde{\mu}_0 = \min_S \left(\frac{\sum_{i \in S} w_i \bar{X}_i}{\sum_{i \in S} w_i} \right) = \text{RMLE}; & \text{with probability } 1 - p = \frac{1}{k+1} \end{cases} \tag{3.3}$$

This means, we give a chance to the control group mean by a randomized decision and since the control mean in each step with probability p select the UMLE and with probability $1 - p$ select the RMLE. If the control group mean be less than of the means with probability $1 - p$ for control group select the RMLE else, with probability p select the UMLE for control group mean. So, with this technique pressure is removed from the control group. Therefore, this procedure does not have the drawbacks of RMLEs and the RMLE of control group mean will not goes to the infinity as the number of comparisons increases.

It is of interest to find the smallest integer k satisfying in (2.5). We consider the case $n_1 = n_2 = \dots = n_k = n$ and $\mu_1 = \mu_2 = \dots = \mu_k = \mu$. Since $\hat{\mu}_0$ is increasing in μ under the extreme condition $\mu_0 = \mu$, we found by simulation that the k can be as small as 6 provided $n_0 > 3n$. When $n_0 = n$, the

smallest k is 10 as shown in Table 1 of the next section.

4 Simulation study

We conducted a simulation study to evaluate the performance of the proposed estimator with respect to another estimators. In this simulation study we consider the problem of estimating μ_0 under the tree order restriction $\mu_0 \leq \mu_i$, for all $i \geq 1$. A large simulation study was conducted with sample size of n per experimental group and three different patterns of k ($k = 6, 10, 20$), various patterns of means μ_0, \dots, μ_k and three different patterns of σ ($\sigma = 1.5, 2.5, 5.5$). We generated data from the normal distribution with means μ_0, \dots, μ_k and equal population variances σ^2 across experimental groups. Since the results that were obtained for these various patterns and a subset of the simulation results is provided here. The results of the simulation study were based on 10000 simulation runs. In our simulation study we investigated the performance of the MSE for the UMLE, RMLE and proposed estimator (RORE). Although in the any procedure we computed the efficiency for the three estimation procedure. The results of our simulation experiment are summarized in Table 1 that represent MSE of (A) UMLE, (B) RMLE and (C) the proposed estimator (RORE) where the parameter of interest is μ_0 and $\mu_0 \leq \mu_i$, for $i = 1, 2, \dots, k$. The overall conclusions the gain in MSE due to the procedures proposed was substantial for normal means under tree order restriction. An important conclusion from the simulation study was that in most patterns the proposed estimator (RORE) have smaller MSE than the two another estimator (UMLE & RMLE). For example, in case 1 of Table 1 the MSE of RORE is 0.92, the MSE of UMLE and RMLE are 1.25 and 1.12 respectively. Furthermore, the proposed estimator seems to attain the minimum MSE between three estimator when k is large. On the other hand, the MSE of RMLE is increasing with respect to k (asymptotically property). The MSE of UMLE must be equal to $\frac{\sigma^2}{n}$ which simulation results in Table 1 is denoted on this argument. Among the three estimation procedures, the proposed estimator (RORE) appeared to have the most efficiency and we observe that RORE procedure performed better than two estimation procedures. In such situations, the RORE has a most efficiency with respect to other estimators, where relative efficiency of RORE with respect to UMLE is 1.36 and relative efficiency of RMLE with respect to UMLE equal to 0.89. Thus the efficiency of RORE is higher than the RMLE in each row of Table

1 or in any iteration of simulation. In particular, our simulation studies suggest that for normally distributed data the proposed estimator compete very well with the UMLE and RMLE. In most cases, they perform better than both the UMLE and RMLE in the MSE and efficiency criterions. In this case, we find that our estimator very well in comparison with the UMLE and RMLE.

Table 1: Mean Squared Error (MSE) of (A) UMLE, (B) RMLE, (C) RORE.

Patterns of Sigma & k	MSE		
	(A)	(B)	(C)
$\sigma = 1.5, k = 6$	1.25	1.12	0.92
$\sigma = 2.5, k = 6$	1.83	1.31	0.98
$\sigma = 5.5, k = 6$	1.98	1.52	1.08
$\sigma = 1.5, k = 10$	1.29	1.45	0.97
$\sigma = 2.5, k = 10$	1.26	1.59	1.11
$\sigma = 5.5, k = 10$	1.27	1.88	1.14
$\sigma = 1.5, k = 20$	1.27	1.88	1.02
$\sigma = 2.5, k = 20$	1.31	1.95	1.08
$\sigma = 5.5, k = 20$	1.35	2.18	1.15

5 Concluding remarks

To motivate the development of some alternate estimators to the restricted maximum likelihood estimators, we estimated the control group mean parameter, μ_0 , in a simple tree ordering. Under the simple tree order restriction, Lee (1988) proved that the restricted maximum likelihood estimator for μ_0 fails to dominate UMLE in the mean squared error sense as $k \rightarrow \infty$. In this paper we provide a representative sample of our results and in all our simulations we considered the worst case scenario for μ vector, which is the boundary point $(0, 0, \dots, 0)$. When the components of μ are estimated simultaneously, the restricted maximum likelihood estimator is known to perform better than the unrestricted maximum likelihood estimator. However, it depends upon the type of order restriction when one is interested in estimating the individual components of μ . For instance, when $\bar{X}_0, \dots, \bar{X}_k$ are independently normally distributed, then under simple tree order restriction the RMLE dominates the UMLE (Kelly (1989), Lee (1981)). However, in the case of simple tree ordering, as observed by Lee (1988),

the RMLE fails disastrously, especially when k is large. In this article we strengthen and unify the previously known results of Lee (1981, 1988) and Kelly (1989) for the tree order restriction. A new scheme is introduced to construct better estimator for tree order restriction. Extensive simulation studies suggest that the proposed estimator, RORE, that is significantly better than the RMLE. As expected, in each case as the dimension increased, the RMLE performed poorly in terms of MSE. In several cases the MSE of RMLE was substantially larger than all its competitors. As expected, the proposed estimator (RORE) performed extremely well when variance of populations was large. However, its performance can be poor for RMLE. The proposed estimator (RORE) consistently performed well for all patterns considered in our simulation studies. Relative to RMLE, in some cases with k is higher than 50, its MSE was almost one-fourth that of RMLE. Therefore, the proposed estimator RORE, performed better than the RMLE in terms of MSE. In conclusion, RORE never failed the way UMLE and was always better than the RMLE.

It will also be interesting to generalize the results of this paper to the similar phenomenon in terms of the hypothesis testing in basis of proposed estimator RORE.

References

- Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical inference under order restrictions: the theory and applications of isotonic regression*, Wiley, New York.
- Kelly, R. (1989). Stochastic reduction of loss in estimating normal means by isotonic regression, *Ann. Statist.*, **17**, 937–940.
- Lee, C. I. C. (1988). The quadratic loss of order restricted estimators for several treatment means with a control mean, *Ann. Statist.*, **16**, 751–758.
- Lee, C.I. C. (1981). The quadratic loss of isotonic regression under normality, *Ann. Statist.*, **11**, 467–477.
- Robertson, T. Wright, F. T., Dykstra, R. L. (1988). *Order restricted statistical inference*, New York, Wiley.

Silvapulle, M. and Sen, P. K. (2005). *Constrained statistical inference: inequality, order and shape restrictions*, Wiley, New York, NY.



Asymptotic behavior of least square estimator for stochastic differential equation arisen from RL electrical circuits

Parisa Nabati ⁸

Urmia University of Technology, Urmia, Iran.

Abstract: This paper investigates the problem of the parameter estimation of stochastic differential equation arisen from RL electrical circuits when the voltage is affected by the noise and the current is observed at the discrete time points. The least square method that is used to estimate the resistor is called \hat{R}_n . The strong consistency of \hat{R}_n is discussed when inductor is a known parameter.

Keywords: Estimation, Least square estimator, Strong consistency, Stochastic differential equations.

Mathematics Subject Classification (2010): 65C05.

1 Introduction

Recently, Stochastic Differential Equations (SDEs) have been utilized to model the systems that are subject to fluctuations. Diffusion processes have been used in modeling various phenomena, for example, noisy electrical circuits, tumor growth and interest rates. The accurate parameter estimation of both the drift and diffusion terms of the aforementioned linear and nonlinear models is a concern. This paper investigates the parameter estimation of the RL electrical circuits which have a different form of SDEs compared to the models that have been studied until now. Any electrical circuit consists of resistor (R), capacitor (C) and inductor (L). These circuit elements can be combined to form an electrical circuit in four distinct way: the RC, RL, LC and RLC circuits. An inductor-resistor circuit (RL), is an electrical circuit composed of resistor and inductor where the power is being fed through a voltage or current source. The Ordinary Differential Equation (ODE) describing the behavior of the RL circuit is given by kirchhoff's current law in kolarova (2005),

$$L \frac{dI}{dt} + RI(t) = V(t), \quad I(0) = I_0, \quad (1.1)$$

⁸Parisa Nabati: p.nabati@uut.ac.ir

where $V(t)$ denotes the potential source at time t .

$(\Omega, \mathcal{F}, \mathbb{P})$ assumed to be a basic probability space equipped with a right continuous family of σ -algebras $\{\mathcal{F}_t, t \geq 0\}$ and $\{\mathcal{W}_t, t \geq 0\}$ is an \mathcal{F}_t adapted Wiener process. If there is some randomness in the potential source, then voltage may not be deterministic but of the form:

$$V^*(t) = V(t) + \epsilon \mathcal{W}(t), \quad (1.2)$$

where ϵ is nonnegative constant, known as the intensity of noise. By substituting this from the equation of the circuit, the SDE will be derived as:

$$dI(t) = \left(\frac{1}{L}V(t) - \frac{R}{L}I(t)\right)dt + \frac{\epsilon}{L}dB(t), \quad (1.3)$$

where $B(t)$ is one dimensional Brownian motion.

Eq. (1.3) is apparently similar to the mean reverting Ornstein-Uhlenbeck process $\{X_t, t \geq 0\}$, starting from $x \in \mathbb{R}$ which is defined as the unique solution to the following linear SDE,

$$dX_t = (a - \theta X_t)dt + \epsilon dB_t,$$

where a, θ and ϵ are given constants. The efficient estimation of the drift parameters of small diffusins from the discrete observations is given in Catalot (1990) and Laredo (1990). Martingale estimating function was used by Sorensen (2000), to establish consistency and asymptotic normality of the estimators of the drift and diffusion coefficient parameters when $\epsilon \rightarrow 0$.

The parameter estimation of Eq. (1.3) has not been studied before. In Eq. (1.3) L is a nonlinear parameter, therefore the problem of parameter estimation in the present work has fundamental different in comparison with the problem of parameter estimation for θ in the O-U process.

Although paths of diffusion processes are continuous, in almost all cases they only could be observed at discrete time moments. Assume that the Eq. (1.3) is observed at regularly spaced time points $\{t_i = \frac{i}{n}, i = 1, 2, \dots, n\}$. It is important to accurately estimate the parameters R and L from the observed data. This leads us to the mathematical problem of estimating the true values of (R, L) from the observation $\{I_{t_i}, i = 1, 2, \dots, n\}$. In order to simplify the mathematical equations, we make the following assumption $t_i = ih, h = \frac{1}{n}$.

Maximum Likelihood Estimator (MLE) is a popular method in the case of diffusion processes driven by

Brownian motions, when the process can be observed continuously, Prakasa (1983). When a diffusion process is observed at discrete times, in most cases the transition density and hence likelihood function of the observations are not explicitly computable. In order to circumvent this concern, some approximate likelihood methods have been proposed by lo (1988) and Pedersen (1983). MLE is asymptotically equivalent to the Least Square Estimator (LSE). The strong consistency of the LSE is studied in kasonga (1988). The convergence in probability is discussed in Dorogovcev (1976) and Breton (1976), and the asymptotic distribution is studied in Prakasa (1983). Singularity of the LSE for mean reverting α -stable motions was studied in Long (2009). The strong consistency and asymptotic behavior of LSE for discretely observed O-U processes with small levy noises is studied in Long (2009). This paper focuses on the strong consistency of the LSE for RL electrical circuits satisfying the Eq. (1.3). The paper is organized as follows. In section 2, the strong consistency of the LSE $\hat{R}_{n,\epsilon}$ will be studied when L is a known parameter. In section 3, a simulation is conducted to verify the accuracy of the proposed method. Finally, some concluding remarks are given in section 4.

2 Strong Consistency

Let $h > 0$ be fixed and let the process $I(t)$ defined by Eq. (1.3) be observed at $t_i = ih$. The LSE of R for Eq. (1.3) is to minimize the following contrast function:

$$\Lambda_n(R) = \Lambda_n(R, \{I_{t_i}\}_{i=0}^n) = \sum_{i=1}^n |L(I_{t_i} - I_{t_{i-1}}) + RI_{t_{i-1}}\Delta t_{i-1} - V\Delta t_{i-1}|^2, \quad (2.1)$$

where $\Delta t_{i-1} = t_i - t_{i-1} = \frac{1}{n}$. Then the LSE $\hat{R}_{n,\epsilon}$ is defined as

$$\hat{R}_{n,\epsilon} = \operatorname{argmin} \Lambda_{n,\epsilon}(R),$$

which can be represented as:

$$\hat{R}_{n,\epsilon} = -L \frac{\sum_{i=1}^n I_{t_{i-1}}(I_{t_i} - I_{t_{i-1}})}{h \sum_{i=1}^n I_{t_{i-1}}^2} + V \frac{\sum_{i=1}^n I_{t_{i-1}}}{\sum_{i=1}^n I_{t_{i-1}}^2}. \quad (2.2)$$

The explicit solution of the Eq. (1.3) is,

$$I_t = e^{-\frac{R}{L}t} \cdot I_0 + \frac{V}{L} \int_0^t e^{-\frac{R}{L}(t-s)} ds + \frac{\epsilon}{L} \int_0^t e^{-\frac{R}{L}(t-s)} dB_s,$$

hence the LSE $\hat{R}_{n,\epsilon}$ can be represented as:

$$\begin{aligned} \hat{R}_{n,\epsilon} &= \frac{L}{h} \left(1 - e^{-\frac{R}{L} \cdot \frac{1}{n}}\right) - \frac{\epsilon}{h} \cdot \frac{\sum_{i=1}^n I_{t_{i-1}} \cdot \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} dB_s}{\sum_{i=1}^n I_{t_{i-1}}^2} + V \frac{\sum_{i=1}^n I_{t_{i-1}}}{\sum_{i=1}^n I_{t_{i-1}}^2} \\ &\quad - \frac{V \sum_{i=1}^n I_{t_{i-1}} \cdot \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} ds}{h \sum_{i=1}^n I_{t_{i-1}}^2}. \end{aligned} \tag{2.3}$$

The asymptotic behavior of the LSE $\hat{R}_{n,\epsilon}$ with high frequency $n \rightarrow \infty$ and small dispersion $\epsilon \rightarrow 0$ will be considered. The goal is to prove that $\hat{R}_{n,\epsilon} \rightarrow R$ in probability.

Theorem 2.1. *Suppose that $h \rightarrow 0$ and $t_n = nh \rightarrow \infty$ as $n \rightarrow \infty$. Then the following strong consistency holds: $\hat{R}_{n,\epsilon} \rightarrow_p R$*

We have stabilized some preliminary results before proving theorem 2.1. Eq. (2.3) can be written as:

$$\begin{aligned} \hat{R}_{n,\epsilon} - R &= [L \cdot n \cdot (1 - e^{-\frac{R}{L} \cdot \frac{1}{n}}) - R] \\ &\quad - \frac{\epsilon \sum_{i=1}^n I_{t_{i-1}} \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} dB_s}{h \sum_{i=1}^n I_{t_{i-1}}^2} \\ &\quad + \frac{V \cdot h \sum_{i=1}^n I_{t_{i-1}} - V \sum_{i=1}^n I_{t_{i-1}} \cdot \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} ds}{h \sum_{i=1}^n I_{t_{i-1}}^2} \\ &:= \Phi_n - \frac{\Psi_1(n,\epsilon)}{\Psi_2(n,\epsilon)} + \frac{\Psi_3(n,\epsilon)}{\Psi_2(n,\epsilon)}. \end{aligned} \tag{2.4}$$

It is clear that $\Phi_n \rightarrow 0$ as $n \rightarrow \infty$. The asymptotic behavior of the $\Psi_1(n, \epsilon), \Psi_2(n, \epsilon)$ and $\Psi_3(n, \epsilon)$ is required to be studied as follows:

Corollary 2.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a basic probability space equipped with a right continuous family of σ -algebras $\{\mathcal{F}_t, t \geq 0\}$ and f_t is \mathcal{F}_t -adapted function, and let $0 \leq S < T$. Then*

$$E\left[\int_S^T f(t) dB_t\right] = 0. \tag{2.5}$$

Proof. See Oksendal (1998). □

Lemma 2.3. *If $n \rightarrow \infty$ and $\epsilon \rightarrow 0$, then $\Psi_1(n, \epsilon) \rightarrow_p 0$.*

Proof. Note that

$$I_{t_{i-1}} = e^{-\frac{R}{L} t_{i-1}} \cdot I_0 + \frac{V}{L} \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} ds + \frac{\epsilon}{L} \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s. \tag{2.6}$$

Then $\Psi_1(n, \epsilon)$ can be written as:

$$\begin{aligned}\Psi_1(n, \epsilon) &= \epsilon(I_0 - \frac{V}{R}) \sum_{i=1}^n e^{-\frac{R}{L}t_{i-1}} \cdot \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} dB_s \\ &+ \epsilon \frac{V}{R} \cdot \sum_{i=1}^n \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} dB_s \\ &+ \frac{\epsilon}{L} \cdot \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s \cdot \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} dB_s \\ &:= \sum_{k=1}^3 \Psi_{1,k}(n, \epsilon).\end{aligned}\tag{2.7}$$

Now, the convergence of the $\Psi_{1,k}(n, \epsilon)$ for $k=1, 2, 3$ should be studied.

by chebyshev's inequality, for any given $\delta > 0$, we have:

$$P(|\Psi_{1,1}(n, \epsilon)| > \delta) \leq \frac{E|\Psi_{1,1}(n, \epsilon)|^2}{\delta^2} = \delta^{-2} (I_0 - \frac{V}{R})^2 \epsilon^2 \cdot \frac{e^{\frac{2R}{Ln}} - 1}{\frac{2R}{L}} \cdot \sum_{i=1}^n e^{-\frac{2R}{L}t_i},\tag{2.8}$$

which tends to zero as $n \rightarrow \infty$ and $\epsilon \rightarrow 0$.

$$\begin{aligned}P(|\Psi_{1,2}(n, \epsilon)| > \delta) &\leq \delta^{-2} E|\epsilon \cdot \frac{V}{R} \cdot \sum_{i=1}^n \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} dB_s|^2 \\ &= \delta^{-2} \cdot \epsilon^2 \cdot \frac{V^2}{R^2} \cdot \frac{L}{2R} \cdot \frac{1-e^{-\frac{2R}{L}h}}{h},\end{aligned}\tag{2.9}$$

which tends to zero as $n \rightarrow \infty$ and $\epsilon \rightarrow 0$.

$$\begin{aligned}P(|\Psi_{1,3}(n, \epsilon)| > \delta) &\leq \delta^{-2} E|\frac{\epsilon}{L} \cdot \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s \cdot \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} dB_s|^2 \\ &= \delta^{-2} \frac{\epsilon^2}{L^2} \cdot \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{2R}{L}(t_{i-1}-s)} ds \cdot \int_{t_{i-1}}^{t_i} e^{-\frac{2R}{L}(t_i-s)} ds \\ &= \delta^{-2} \frac{\epsilon^2}{L^2} \cdot \frac{1-e^{-\frac{2R}{L} \cdot \frac{1}{n}}}{\frac{2R}{L}} \cdot \sum_{i=1}^n e^{-\frac{2R}{L}t_i},\end{aligned}\tag{2.10}$$

which tends to zero as $n \rightarrow \infty$ and $\epsilon \rightarrow 0$.

Hence it can be concluded that $\Psi_1(n, \epsilon) \rightarrow_P 0$. \square

Lemma 2.4. Show that $\Psi_2(n, \epsilon) \rightarrow_P I_0^2 \int_0^1 e^{-\frac{2R}{L}s} ds + \frac{V^2}{L^2} \cdot \frac{1}{2R} (1 - \int_0^1 e^{-\frac{2R}{L}s} ds)$. as $n \rightarrow \infty$ and $\epsilon \rightarrow 0$.

Proof.

$$\begin{aligned}\Psi_2(n, \epsilon) &= \frac{1}{n} \cdot \sum_{i=1}^n I_{t_{i-1}}^2 \\ &= \frac{1}{n} \cdot \sum_{i=1}^n (I_0 e^{-\frac{R}{L}t_{i-1}} + \frac{V}{L} \cdot \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} ds + \frac{\epsilon}{L} \cdot \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s)^2 \\ &= \frac{1}{n} \sum_{i=1}^n I_0^2 e^{-\frac{2R}{L}t_{i-1}} \\ &+ \frac{1}{n} \cdot \frac{V^2}{L^2} \cdot \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{2R}{L}(t_{i-1}-s)} ds \\ &+ \frac{1}{n} \cdot \frac{\epsilon^2}{L^2} \cdot \sum_{i=1}^n (\int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s)^2 \\ &+ \frac{2}{n} \cdot \frac{V}{L} I_0 \cdot \sum_{i=1}^n e^{-\frac{R}{L}t_{i-1}} \cdot \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} ds \\ &+ \frac{2}{n} \cdot \frac{\epsilon}{L} I_0 \cdot \sum_{i=1}^n e^{-\frac{R}{L}t_{i-1}} \cdot \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s \\ &+ \frac{2}{n} \cdot \frac{V}{L} \cdot \frac{\epsilon}{L} \cdot \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} ds \cdot \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s \\ &:= \sum_{k=1}^6 \Psi_{2,k}(n, \epsilon).\end{aligned}\tag{2.11}$$

It is clear that:

$$\Psi_{2,1}(n, \epsilon) = I_0^2 \cdot \frac{1}{n} \sum_{i=1}^n e^{-\frac{2R}{L}t_{i-1}} \rightarrow I_0^2 \int_0^1 e^{-\frac{2R}{L}s} ds, \quad (2.12)$$

as $n \rightarrow \infty$.

for

$$\Psi_{2,2}(n, \epsilon) = \frac{1}{n} \cdot \frac{V^2}{L^2} \cdot \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{2R}{L}(t_{i-1}-s)} ds \rightarrow \frac{V^2}{L^2} \cdot \frac{1}{2R} (1 - \int_0^1 e^{-\frac{2R}{L}s} ds), \quad (2.13)$$

as $n \rightarrow \infty$.

For $\Psi_{2,3}(n, \epsilon)$ by Markov's inequality, the following Eq. can be derived:

$$\begin{aligned} P(|\Psi_{2,3}(n, \epsilon)| > \delta) &\leq \delta^{-1} E \left| \frac{1}{n} \cdot \frac{\epsilon^2}{L^2} \cdot \sum_{i=1}^n \left(\int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s \right)^2 \right| \\ &= \delta^{-1} \cdot \frac{\epsilon^2}{L^2} \cdot \frac{1}{n} \cdot \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{2R}{L}(t_{i-1}-s)} ds \\ &= \delta^{-1} \cdot \frac{\epsilon^2}{L^2} \cdot \frac{1}{n} \cdot \sum_{i=1}^n e^{-\frac{2R}{L}t_{i-1}} \cdot \int_0^{t_{i-1}} e^{\frac{2R}{L}s} ds, \end{aligned} \quad (2.14)$$

which tends to zero as $n \rightarrow \infty$. Eq. (2.14) describes the $\Psi_{2,4}(n, \epsilon)$:

$$\Psi_{2,4}(n, \epsilon) = \frac{2}{n} \cdot \frac{V}{R} I_0 \cdot \sum_{i=1}^n e^{-\frac{R}{L}t_{i-1}} (1 - e^{-\frac{R}{L}t_{i-1}}), \quad (2.15)$$

which tends to zero when $n \rightarrow \infty$. For $\Psi_{2,5}(n, \epsilon)$ and $\Psi_{2,6}(n, \epsilon)$ similar to the proof of equation (2.14), we get

$$P(|\Psi_{2,5}(n, \epsilon)| > \delta) \leq \delta^{-2} \cdot \left(\frac{2}{n} \cdot \frac{\epsilon}{L} I_0 \right)^2 \cdot \frac{e^{\frac{2R}{L}-1}}{\frac{2R}{L}} \cdot \sum_{i=1}^n e^{-\frac{2R}{L}t_i}, \quad (2.16)$$

which tends to zero and

$$P(|\Psi_{2,6}(n, \epsilon)| > \delta) \leq \delta^{-2} \cdot \left(\frac{V \cdot \epsilon}{n \cdot L \cdot R} \right)^2 \cdot \sum_{i=0}^n (1 - e^{-\frac{R}{L}t_{i-1}})^2 \quad (2.17)$$

which tends to zero. Therefore, by combining 2.11-2.17, we conclude that lemma 2.4 holds. \square

Lemma 2.5. *Show that $\Psi_3(n, \epsilon)$ tends to zero as $n \rightarrow \infty$ and $\epsilon \rightarrow 0$.*

Proof.

$$\begin{aligned} \Psi_3(n, \epsilon) &= h \sum_{i=1}^n I_{t_{i-1}} - \sum_{i=1}^n I_{t_{i-1}} \cdot \int_{t_{i-1}}^{t_i} e^{-\frac{R}{L}(t_i-s)} ds \\ &= \left[h - \frac{L}{R} (1 - e^{-\frac{R}{L}h}) \right] \cdot \sum_{i=1}^n I_{t_{i-1}} \\ &= \left[h - \frac{L}{R} (1 - e^{-\frac{R}{L}h}) \right] \sum_{i=1}^n \left(e^{-\frac{R}{L}t_{i-1}} \cdot I_0 + \frac{V}{L} \int_0^{t_i} e^{-\frac{R}{L}(t_{i-1}-s)} ds \right) \\ &\quad + \frac{\epsilon}{L} \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} dB_s \\ &:= \sum_{k=1}^3 \Psi_{3,k}(n, \epsilon). \end{aligned} \quad (2.18)$$

For $\Psi_{3,1}(n, \epsilon)$, it is clear that,

$$\Psi_{3,1}(n, \epsilon) = I_0 \left[h - \frac{L}{R} (1 - e^{-\frac{R}{L}h}) \right] \sum_{i=1}^n e^{-\frac{R}{L}t_{i-1}} \rightarrow I_0 \left[1 - \frac{L}{Rh} (1 - e^{-\frac{R}{L}h}) \right] \cdot \int_0^1 e^{-\frac{R}{L}s} ds, \quad (2.19)$$

which tends to zero as $h \rightarrow 0$. For $\Psi_{3,2}(n, \epsilon)$ we have,

$$\begin{aligned} \Psi_{3,2}(n, \epsilon) &= \frac{V}{L} \left[h - \frac{L}{R} (1 - e^{-\frac{R}{L}h}) \right] \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{R}{L}(t_{i-1}-s)} ds \\ &\rightarrow \frac{V}{R} \left[1 - \frac{L}{Rh} (1 - e^{-\frac{R}{L}h}) \right] \cdot (1 - \int_0^1 e^{-\frac{R}{L}s} ds), \end{aligned} \quad (2.20)$$

which tends to zero as $h \rightarrow 0$. □

For $\Psi_{3,3}(n, \epsilon)$ by chebysheve's inequality for any given $\delta > 0$,

$$P(|\Psi_{3,3}(n, \epsilon)| > \delta) \leq \delta^{-2} \cdot (h - \frac{L}{R} (1 - e^{-\frac{R}{L}h}))^2 \cdot \frac{\epsilon^2}{L^2} \cdot \sum_{i=1}^n \int_0^{t_{i-1}} e^{-\frac{2R}{L}(t_{i-1}-s)} ds. \quad (2.21)$$

which tends to zero as $h \rightarrow 0$ and $\epsilon \rightarrow 0$. Then lemma 2.5 holds.

Proof. Now we can proof the theorem 2.1. By using the fact that $\Phi_n \rightarrow 0$ as $n \rightarrow \infty$, and combining lemma 2.3, 2.4 and 2.5 it can be concluded that $\hat{R}_{n,\epsilon} \rightarrow_p R$. □

3 Numerical Simulation

The RL electrical circuit with the unknown parameter R is considered for the first simulation. The objective is to estimate the parameter R using the least square estimator based on the discrete observation $\{I_{t_i}\}_{i=1}^n$. In the simulation, $R = 15$ and use values of n and ϵ vary. The results are shown in Table 1. From Table 1, it is can be claimed that when $\epsilon \rightarrow 0$, the obtained estimation values are very close to the true parameter value $R = 15$.

Table 1: The mean absolute deviation of the estimator \hat{R}_n with different noise intensity.

n	$\epsilon = 10$	$\epsilon = 1$	$\epsilon = 0.1$	$\epsilon = 0.01$	$\epsilon = 0.001$
100	0.3139	0.0144	2.3506×10^{-4}	5.6253×10^{-5}	2.5983×10^{-6}
1000	0.3116	0.0055	1.0998×10^{-4}	2.3030×10^{-5}	1.6457×10^{-6}
10000	0.2827	0.0026	3.3854×10^{-5}	4.6923×10^{-6}	1.1653×10^{-6}

4 Conclusion

This paper proposes the Least Square Estimator for the parameter R of the RL electrical circuits when L is a known parameter. Some new asymptotic results are stabilized in order to show that this estimator converges in probability. It should be mentioned that L is a nonlinear parameter and then this makes changes the problem of parameter estimation of present work in comparison with O-U process which is considered in Catalot (1990) and Laredo (1990).

References

- Breton, A. L. (1976). On continuous and discrete sampling for parameter estimation in diffusion type processes, *Math program.*
- Dorogovcev, A. Ja. (1976). The consistency of an estimate of a parameter of a stochastic differential equation, *Theory probab.math.statist.*
- Catalot, V. G. (1990) Maximum contrast estimation for diffusion processes from discrete observations, *Statistics.*
- Kasonga, R. A. (1988). The consistency of a nonlinear least square estimator for diffusion processes, *Stochastic processes Appl.*
- Kolarova, E. (2005). Modeling RL electrical circuits by SDEs, *International conference on computer as a tool.*
- Laredo, C. F. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process, *Ann. Statist.*
- lo, A. W. (1988). Maximum likelihood estimation of generalized Ito processes with discretely sampled data, *Economic Theory 4.*
- Long, H. (2009). Least square estimator for discretely observed Ornstein-Uhlenbeck processes with small levy noises, *Statistics and probability letters.*

Oksendal, B. (1998). Stochastic Differential Equations, *Springer*, Fifth edition.

Pedersen, A.R. (1995). A new approach to Maximum likelihood estimation for stochastic differential equations based on discrete observations, *scand. J. Statist.* 22.

Prakasa Rao, B. L. S. (1983). Asymptotic theory for nonlinear least square estimator for diffusion processes, *Math.oper.forsch.stat.*

Sorensen, M. (2000). Small dispersion asymptotics for diffusion martingale estimating functions, Preprint No. 2000-2, Department of statistics and operation Research, University of Copenhagen.



Complete Convergence for Weighted sums of FGM random sequences

Naderi, H.¹, Amini, M.² Bozorgnia, A.³

^{1,2,3}Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract: Some complete convergence theorems for weighted sums of FGM random sequences are provided and applied to empirical distribution, sample p th quantile and sample mean of random weighting estimate.

Keywords Complete convergence, FGM sequence, Copula.

MSC(2010):60F15

1 Introduction

Hsu and Robbins (1947) proved the sequence of arithmetic means of i.i.d. random variables converges completely to the expected value if the variance of the summands is finite. Erdős (1949) proved the converse. The result of Hsu-Robbins-Erdős is a fundamental theorem in probability theory and has been generalized and extended in several directions by many authors. Recently, Amini et al.(2012) studied complete convergence of moving-average processes under negative dependent Sub-Gaussian assumptions, Amini et al. (2015) investigated complete convergence of moving-average processes based on an identically distributed doubly infinite sequence of negatively superadditive-dependent random variables, Naderi et al.(2015) derived Complete Convergence for Weighted Sums of Weakly Negative Dependent of Random Variables, Qiu et al. (2013) obtained the complete convergence for arrays of rowwise extended negatively dependent random variables, Wang et al. (2014) discussed the complete convergence for arrays of rowwise NSD random variables and its applications, Shen et al. (2012) studied almost sure convergence and strong stability for weighted sums of NSD random variables,

²Amini. M: bastani@iasbs.ac.ir

also Sung (2012), Wang, et, al.(2012),Wang,et,al.(2014), Wu (2012) and Shen et al. (2011). In this paper, we prove complete convergence for weighted sums of FGM sequences under certain dependence structures. The paper is organized as follows. Some useful definitions, basic properties are given in Section 2. The main results are presented in Section 3.

2 Definition and properties

This section contains main definitions and some properties of sequences random variables with certain dependence structures.

A copula is a function that links univariate marginal distributions to the full multivariate distribution and it's a general tool for assessing the dependence structure of random variables. For details see Nelsen (2006). An n -dimensional copula is a function C with domain $[0, 1]^n$ if

- (i) $C(\mathbf{u}) = 0$ whenever $\mathbf{u} \in [0, 1]^n$ has at least one component equal to 0,
- (ii) $C(\mathbf{u}) = u_i$ whenever $\mathbf{u} \in [0, 1]^n$ has all components equal to 1 except the i -th one, which is equal to u_i ,
- (iii) C is n -increasing,i.e., for each n -box $B = \times_{i=1}^n [u_i, v_i]$ in $[0, 1]^n$ with $u_i \leq v_i$ for each $i \in 1, \dots, n$,

$$V_C(B) = \sum_{\mathbf{z} \in \times_{i=1}^n [u_i, v_i]} (-1)^{N(\mathbf{z})} C(\mathbf{z}) \geq 0,$$

where $N(\mathbf{z}) = \text{card}\{k | z_k = u_k\}$.

More formally, Sklar's Theorem (1959) states that any continuous the joint distribution can be uniquely represented by a n -copula

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)),$$

where F is the joint distribution to the marginal distribution functions $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$. It is easy to construct the corresponding copula as;

$$C(u_1, u_2, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)), \quad \text{for all } u_i \in [0, 1], i = 1, 2, \dots, n.$$

The copula density c is the derivative of C with respect to each of its arguments as:

$$c(u_1, u_2, \dots, u_n) = \frac{\partial^n}{\partial u_1 \dots \partial u_n} C(u_1, \dots, u_n).$$

By using the chain rule, the density function $f(x_1, x_2, \dots, x_n)$ corresponding to the copula $C(x_1, x_2, \dots, x_n)$ is,

$$f(x_1, x_2, \dots, x_n) = c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i)$$

where $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$ are the marginal densities of f . Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector having to joint distribution function $F(x_1, \dots, x_n)$ and absolute continuous marginals F_1, \dots, F_n , with corresponding copula function $C(u_1, \dots, u_n)$.

If $c(u_1, \dots, u_n) \leq M_1$ for some $M_1 \geq 1$, then \mathbf{X} having to negative dependence structure. (The FGM n-copulas). Let

$$C(u_1, u_2, \dots, u_n) = \prod_{i=1}^n u_i \left(1 + \sum_{1 \leq j < k \leq n} \alpha_{kj} \bar{u}_j \bar{u}_k \right) \quad (|\alpha_{kj}| \leq 1).$$

where $|\alpha_{kj}| \leq 1$ and $1 + \sum_{1 \leq k < j \leq n} \alpha_{kj} \geq 0$. Also, $\bar{u}_{jk} = 1 - F_{jk}$ for $k = 2, \dots, n$.

Via Sklar's Theorem, it's easy to show that, the copula density is

$$c(u_1, u_2, \dots, u_n) = 1 + \sum_{1 \leq j < k \leq n} \alpha_{kj} (1 - 2u_j)(1 - 2u_k).$$

It is easy to show that $c(u_1, \dots, u_n) \leq 1 + \sum_{s=2}^n \sum_{1 \leq j < k \leq n} |\alpha_{kj}|$, so there exist $M_1 \geq 1$ such that Lemma 2.2 holds for $-1 < \alpha_{kj} < 0$, for all j, k .

The joint distributions with given margins F_1, F_2, \dots, F_n generated by this copula having to negative dependence structure and also,

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_i(x_i) \left(1 + \sum_{1 \leq k < j \leq n} \alpha_{kj} (1 - F_k(x_k))(1 - F_j(x_j)) \right),$$

Another form of FGM n -copulas defined as;

$$C(u_1, u_2, \dots, u_n) = \prod_{i=1}^n u_i \left(1 + \sum_{l=2}^n \sum_{1 \leq j_1 < \dots < j_l \leq n} \theta_{j_1 j_2 \dots j_l} (1 - u_{j_1})(1 - u_{j_2}) \dots (1 - u_{j_l}) \right).$$

Where $-1 \leq \theta_{j_1, \dots, j_l} \leq 1$ for all j_1, \dots, j_l .

$$c(u_1, u_2, \dots, u_n) = 1 + \sum_{r=2}^n \sum_{1 \leq j_1 < \dots < j_r \leq n} \theta_{j_1 j_2 \dots j_r} (1 - 2u_{j_1})(1 - 2u_{j_2}) \dots (1 - 2u_{j_r}).$$

It is easy to show that $c(u_1, u_2, \dots, u_n) \leq 1 + \sum_{s=2}^n \sum_{1 \leq j_1 < \dots < j_s \leq n} |\theta_{j_1 j_2 \dots j_s}|$, so there exist $M_1 \geq 1$ such that Lemma 2.2 holds for $-1 < \theta_{j_1 j_2 \dots j_l} < 0$.

The joint distributions with given margins F_1, F_2, \dots, F_n generated by this copula having to negative dependence structure and also,

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_i(x_i) \left(1 + \sum_{l=2}^n \sum_{1 \leq j_1 < \dots < j_l \leq n} \theta_{j_1 j_2 \dots j_l} \bar{F}_{j_1}(x_{j_1}) \bar{F}_{j_2}(x_{j_2}) \dots \bar{F}_{j_l}(x_{j_l}) \right).$$

Where $-1 \leq \theta_{j_1, \dots, j_l} \leq 1$ for all j_1, \dots, j_l . (Cambanis, 1993). A sequence $\{X_n, n \geq 1\}$ is called a FGM random sequence, if for all $n \geq 1$ and $x_1, x_2, \dots, x_n \in R$,

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_i(x_i) \left(1 + \sum_{1 \leq k < j \leq n} \alpha_{kj} (1 - F_k(x_k))(1 - F_j(x_j)) \right),$$

where $|\alpha_{kj}| \leq 1$ and $1 + \sum_{1 \leq k < j \leq n} \alpha_{kj} \geq 0$. (Ranjbar et al. 2009) The random vector $\mathbf{X} = (X_1, \dots, X_n)$ is said to be weakly negative dependent (WND) if there exist a constant $M_1 \geq 1$, such that for $n > 1$

$$f(x_1, \dots, x_n) \leq M_1 \prod_{i=1}^n f_i(x_i).$$

Where $f(x_1, \dots, x_n)$ and $f_i(x_i)$'s are joint density and marginal densities of random vector \mathbf{X} , respectively.

We say $f(x_1, \dots, x_n)$ is WND, if satisfy in Definition 2.6. A sequence $\{X_i, i \geq 1\}$ of random variables is said to be pairwise WND if X_i and X_j are WND for all $i \neq j$ and also, is said to be WND if every finite subset X_1, \dots, X_n is WND.

(Liu, 2009) The random vector $\mathbf{X} = (X_1, \dots, X_n)$ is said to be extended negative dependent (END) if there exist a constant $M_2 > 0$ such that both

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) \leq M_2 \prod_{i=1}^n P(X_i \leq x_i)$$

and

$$P(X_1 > x_1, \dots, X_n > x_n) \leq M_2 \prod_{i=1}^n P(X_i > x_i)$$

holds for $n \geq 1$ and $x_1, \dots, x_n \in \mathcal{R}$.

It is obvious that the sequence $\{X_i, i \geq 1\}$ is negative dependent (ND) if $M_2 = 1$.

(i) Let $\{X_n, n \geq 1\}$ be a FGM random sequence, then for all $n \geq 2$, then (X_1, X_2, \dots, X_n) is WND with $M_1 = 1 + \beta$ where $\beta = \sum_{1 \leq i < j \leq n} |\alpha_{ij}|$.

(ii) If $\mathbf{X} = (X_1, \dots, X_n)$ be a WND random vector then is END with $M_2 \geq 1$. In fact the WND condition presented in Definition 2.6 is a useful criterion for characterization class of END random variables. We list some basic properties of FGM random sequences based on our results.

Let $\{X_n, n \geq 1\}$ be a FGM random sequence, then

P_1 . If $\{h_n(x), n \geq 1\}$ is a sequence of increasing real functions then $\{h_n(X_n), n \geq 1\}$ is a FGM random sequences.

P_2 . If $h_1(\cdot), \dots, h_n(\cdot)$ are non-negative real function and

$-1 < \alpha_{kj} < 0$, for all j, k , then for all $n \geq 2$,

$$E\left(\prod_{i=1}^n h_i(X_i)\right) \leq M_1 \prod_{i=1}^n E(h_i(X_i)).$$

In particular, if $h_i(x_i) = e^{tx_i}$, for all $i \geq 1$ and some real t , then;

$$E\left(\prod_{i=1}^n e^{tX_i}\right) \leq M_1 \prod_{i=1}^n E(e^{tX_i}).$$

P_3 . For all $i \neq j$, by Hoeffding Lemma (1963),

$$\text{Cov}(X_i, X_j) = \alpha_{ij} \int_{\mathbb{R}} F_i(x) \cdot \bar{F}_i(x) dx \int_{\mathbb{R}} F_j(y) \bar{F}_j(y) dy.$$

Before we prove the complete convergence for sum and weighted sum of FGM random sequence, we reminder the weak law of large numbers for $\frac{\sum_{i=1}^n X_i}{n}$ proved by Mikami (1997). Let $\{X_n, n \geq 1\}$ be identically FGM random sequence with $X_1 \in L^2$ and $\sum_{1 \leq k < j \leq n} \alpha_{kj} = o(n^2)$, then

$$\frac{\sum_{i=1}^n X_i}{n} \longrightarrow E(X_1), \quad \text{in Probability.}$$

Proof: By Chebychef's inequality for all $\varepsilon > 0$, we get

$$\begin{aligned} P\left[\left|\sum_{i=1}^n (X_i - n \cdot E(X_1))\right| > n \cdot \varepsilon\right] &\leq \frac{1}{n^2 \varepsilon^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n \varepsilon^2} \text{Var}(X_1) \\ &+ \left\{\int_{-\infty}^{+\infty} F_1(x) \cdot \bar{F}_1(x) dx\right\} \frac{1}{n^2} \sum_{1 \leq k < j \leq n} \alpha_{kj}. \end{aligned}$$

This completes the proof.

3 Complete Convergence Theorems

In this section, we derive the complete convergence for weighted sums of FGM random sequence. Also, we discuss its applications in empirical distribution and sample p th quantile. The following theorem concerning the rate of complete convergence of weighted sum $\sum_{i=1}^n a_{ni} X_i$ under some suitable conditions on the coefficients. Let $\{X_i, i \geq 1\}$ be a centered FGM random sequence such that for some $\alpha_i > 0$ and all $t \in \mathbb{R}$, $E\{e^{tX_i}\} \leq e^{\frac{t^2 \alpha_i^2}{2}}$, $i \geq 1$ with $\sum_{i=1}^n \alpha_i^2 = O(n)$. If $\{a_{ni}, i \geq 1, n \geq 1\}$ is an

array of positive constants such that $\sum_{i=1}^n a_{ni}^2 = O(n^{-r})$ for $r > 0$, then for all $\beta > 0$ and $\varepsilon > 0$

$$\sum_{n=1}^{\infty} n^{\beta} P\left(\left|\sum_{i=1}^n a_{ni} X_i\right| > \varepsilon\right) < \infty.$$

Proof: Let $\varepsilon > 0$, $C > 0$ and $A_n = \sum_{i=1}^n a_{ni}^2$. By using P_1 , P_2 and Marov's inequality, we have

$$\begin{aligned} \sum_{n=1}^{\infty} n^{\beta} P\left(\left|\sum_{i=1}^n a_{ni} X_i\right| > \varepsilon\right) &= \sum_{n=1}^{\infty} n^{\beta} P\left(\sum_{i=1}^n a_{ni} X_i > \varepsilon\right) + \sum_{n=1}^{\infty} n^{\beta} P\left(-\sum_{i=1}^n a_{ni} X_i > \varepsilon\right) \\ &\leq 2M_1 \sum_{n=1}^{\infty} n^{\beta} \exp\{-t\varepsilon\} \prod_{i=1}^n E(e^{ta_{ni} X_i}) \\ &\leq 2M_1 \sum_{n=1}^{\infty} n^{\beta} \exp\left\{-t\varepsilon + \frac{\alpha^2 t^2}{2} A_n\right\} \\ &\leq 2M_1 \sum_{n=1}^{\infty} n^{\beta} \exp\left\{-\frac{\varepsilon^2}{2\alpha^2 C} n^r\right\} < \infty. \text{ (by } A_n \leq Cn^{-r}\text{)} \end{aligned}$$

The last inequality holds because, for all $r > 0$ and $\beta > 0$ we have

$$\sum_{n=1}^{\infty} n^{\beta} \exp\left\{-\frac{\varepsilon^2}{2\alpha^2 C} n^r\right\} < \infty \Leftrightarrow \int_1^{\infty} x^{\beta} \exp\left\{-\frac{\varepsilon^2}{2\alpha^2 C} x^r\right\} dx < \infty,$$

this completes the proof.

Let $\{X_i, i \geq 1\}$ be a centered FGM random sequence such that for some $\alpha > 0$ and all $t \in R$, $E\{e^{tX_i}\} \leq e^{\frac{t^2 \alpha^2}{2}}$, then for $p > 1/2$

$$\frac{1}{n^p} \sum_{i=1}^n X_i \rightarrow 0 \text{ completely.}$$

Proof: Let $a_{ni} = 0$ if $i > n$ and $a_{ni} = n^{-p}$ if $i \leq n$. Hence conditions Theorem 3.1 holds for $p > 1/2$. Therefore, by Borel-Cantelli lemma the proof completes. Suppose that $\{X_i, i \geq 1\}$ is a centered FGM random sequence such that $|X_i| < M < \infty$ a.e., ..., $i \geq 1$. Then by Corollary 3.2, for $p > 1/2$

$$\frac{1}{n^p} \sum_{i=1}^n X_i \rightarrow 0 \text{ completely.}$$

As some applications, we study the limiting behavior of the empirical distribution function and sample quantiles of FGM random sequences. Let $\{X_n, n \geq 1\}$ be a FGM random sequence with $F_i(x) = F(x)$, $i \geq 1$.

(i) Suppose that $F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$ is empirical distribution function. Then for every $x \in R$, $\varepsilon > 0$ and $n \geq 1$ we have

$$\sum_{n=1}^{\infty} n^{\beta} P[|F_n(x) - F(x)| > \varepsilon] < \infty \text{ for } \beta > 0.$$

(ii) Let $0 < p < 1$. Suppose that ξ_p is the unique solution x of $F(x^-) \leq p \leq F(x)$. Then for every $\varepsilon > 0$ and $n \geq 1$ we have

$$\sum_{n=1}^{\infty} n^{\beta} P(|\widehat{\xi}_{pn} - \xi_p| > \varepsilon) < \infty \text{ for } \beta > 0.$$

Where $\widehat{\xi}_{pn}$ is the sample p th quantile.

Proof: (i) we define

$$Z_i = I(X_i \leq x), i = 1, \dots, n.$$

The joint density function of (X_1, X_2, \dots, X_n) satisfies in Definition 2.6, so by P_1 the joint density function of (Z_1, Z_2, \dots, Z_n) satisfies in Definition 2.6 also. Since $|Z_i - F(x)| \leq 1$ for $i \geq 1$ then, it is easy show that, for all $t \in R$, $E\{e^{t(Z_i - F(x))}\} \leq e^{\frac{t^2}{2}}$, $i \geq 1$. Thus, for $p = 1$ and $\beta > 0$, we can get

$$\sum_{n=1}^{\infty} n^{\beta} P[|F_n(x) - F(x)| > \varepsilon] < \infty.$$

(ii) For every $\varepsilon > 0$,

$$\begin{aligned} P[|\widehat{\xi}_{pn} - \xi_p| > \varepsilon] &= P[\widehat{\xi}_{pn} > \varepsilon + \xi_p] + P[\widehat{\xi}_{pn} < \xi_p - \varepsilon] \\ &= P[p > F_n(\xi_p + \varepsilon)] + P[p < F_n(\xi_p - \varepsilon)]. \end{aligned}$$

Putting,

$$V_i = I_{[X_i > \xi_p + \varepsilon]} \quad \text{and} \quad U_i = I_{[X_i \leq \xi_p - \varepsilon]}$$

Similarly, by P_1 the joint density functions of random vectors (U_1, U_2, \dots, U_n) and (V_1, V_2, \dots, V_n) satisfy in definition 2.6. Since $|V_i - EV_i| \leq 1$, applying Theorem 3.1, we obtain

$$\begin{aligned} \sum_{n=1}^{\infty} n^{\beta} P[p > F_n(\xi_p + \varepsilon)] &= \sum_{n=1}^{\infty} n^{\beta} P[\sum_{i=1}^n (V_i - E(V_i)) > n\delta_1] \\ &\leq M_1 \sum_{n=1}^{\infty} n^{\beta} \exp\{-\frac{1}{2}n\delta_1^2\} < \infty. \end{aligned}$$

Where $\delta_1 = F(\xi_p + \varepsilon) - p$. Similarly,

$$\begin{aligned} \sum_{n=1}^{\infty} n^{\beta} P[p < F_n(\xi_p - \varepsilon)] &= \sum_{n=1}^{\infty} n^{\beta} P[\sum_{i=1}^n (U_i - E(U_i)) > n\delta_2] \\ &\leq \sum_{n=1}^{\infty} n^{\beta} M_1 \exp\{-\frac{1}{2}n\delta_2^2\} < \infty. \end{aligned}$$

Where $\delta_2 = p - F(\xi_p - \varepsilon)$, these complete the proof.

The following corollary shows that if $\{X_n, n \geq 1\}$ be a FGM random sequence, then $F_n(x)$ and $\widehat{\xi}_{pn}$ are strongly consistent estimators of $F(x)$ and ξ_p respectively. Under the assumptions of Theorem 3.3

(i) For all $x \in R$,

$$F_n(x) \rightarrow F(x) \text{ completely.}$$

(ii) For $0 < p < 1$, then

$$\widehat{\xi}_{pn} \rightarrow \xi_p \text{ completely.}$$

Random weighting is an emerging computing method in statistics, it has been put forward by Zheng.Z (1987). Since this method has many advantages, it drew much attention of statisticians. Gao et al. (2003) studied law of large numbers for sample mean of random weighting estimate for i.i.d random variables. The next theorem determine completely convergence for difference of between random weighting estimate sample mean and sample mean for uniformly bounded FGM random sequence. Let $\{X_n, n \geq 1\}$ be a FGM random sequence with $F_n(x) = F(x)$, $P(a < X_n < b) = 1$, $n \geq 1$).

Suppose that

$T_n = \sum_{i=1}^n \alpha_i X_i$ such that $\sum_{i=1}^n \alpha_i = 1$ for real values $0 \leq a_i \leq 1$. If $\sum_{i=1}^{\infty} \exp(\frac{-\varepsilon^2}{(b-a)B_n}) < \infty$, then

$$T_n - \bar{X}_n = \sum_{i=1}^n \theta_i X_i \rightarrow 0 \text{ completely.}$$

Where $B_n = \sum_{i=1}^n \theta_i^2$ and $\theta_i = \alpha_i - \frac{1}{n}$ and $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$. **Proof:** Taking, $A = \{i, \theta_i \geq 0\}$ and $B = \{i, \theta_i < 0\}$, we can write

$T_n - \bar{X}_n = \sum_{i \in A} \theta_i X_i + \sum_{i \in B} \theta_i X_i$. So, for all $h > 0$, $\varepsilon > 0$ and

$M_1 = 1 + \sum_{1 \leq i < j \leq n} |\alpha_{ij}|$, since for all $h > 0$, $E(e^{hX} \leq \exp \frac{h^2(b-a)^2}{2})$, (Hoeffding Lemma) by applying

Cauchy-Schwarz's inequality and simple computing we have

$$\begin{aligned} P(T_n - \bar{X}_n > \varepsilon) &= P\left(\sum_{i=1}^n \theta_i X_i > \varepsilon\right) \\ &\leq e^{-h\varepsilon} E(e^{h \sum_{i \in A} \theta_i X_i} \cdot e^{h \sum_{i \in B} \theta_i X_i}) \\ &\leq e^{-h\varepsilon} \left\{ E(e^{2h \sum_{i \in A} \theta_i X_i}) \right\}^{\frac{1}{2}} \cdot \left\{ E(e^{2h \sum_{i \in B} \theta_i X_i}) \right\}^{\frac{1}{2}} \quad (C.S.I) \\ &\leq e^{-h\varepsilon} \left\{ M_1 \prod_{i \in A} E(e^{2h\theta_i X_i}) \right\}^{\frac{1}{2}} \cdot \left\{ M_1 \prod_{i \in B} E(e^{2h\theta_i X_i}) \right\}^{\frac{1}{2}}, \quad (P2) \\ &\leq e^{-h\varepsilon} \left\{ M_1 \prod_{i \in A} e^{2h\mu\theta_i + \frac{1}{2}h^2(b-a)^2\theta_i^2} \right\}^{\frac{1}{2}} \cdot \left\{ M_1 \prod_{i \in B} e^{2h\mu\theta_i + \frac{1}{2}h^2(b-a)^2\theta_i^2} \right\}^{\frac{1}{2}} \\ &\leq M_1 \exp\left\{-h\varepsilon + \frac{1}{4}h^2(b-a)^2 B_n\right\}. \end{aligned}$$

Optimizing the exponent, we find $h = \frac{2\varepsilon}{(b-a)^2 B_n}$. So,

$$P(T_n - \bar{X}_n > \varepsilon) \leq M_1 \exp\left\{-\frac{\varepsilon^2}{(b-a)^2 B_n}\right\}.$$

By replacing X_i by $-X_i$ in the above statement, we obtain

$$P(T_n - \bar{X}_n < -\varepsilon) \leq M_1 \exp\left\{-\frac{\varepsilon^2}{(b-a)^2 B_n}\right\}.$$

Observing that

$$\sum_{i=1}^{\infty} P(|T_n - \bar{X}_n| > \varepsilon) \leq 2M_1 \sum_{i=1}^{\infty} \exp\left\{-\frac{\varepsilon^2}{(b-a)^2 B_n}\right\} < \infty.$$

This completes the proof.

The next corollary states results correspondent to the Theorem 3.6 that is related to the limiting distribution of random weighting estimate sample mean that similar to the sample mean and complete convergence of random weighting estimate empirical distribution. Under the assumptions of Theorem 3.6,

(i) If $\bar{X}_n \rightarrow Z$ in distribution, then $T_n \rightarrow Z$ in distribution.

(ii) If $Z_n(x) = \sum_{i=1}^n \alpha_i Z_i(x)$, then for each $x \in \mathcal{R}$,

$$Z_n(x) - F_n(x) \rightarrow 0 \text{ completely.}$$

Where $Z_i(x) = I(X_i \leq x)$.

The next theorem, prove the complete convergence for unbounded FGM random sequence. Let $\{X_i, i \geq 1\}$ be a FGM random sequence with $X_i^2 \in R$ for $i \geq 1$. If $\sum_{i=1}^n \text{Var}(X_i) = O(n^{\delta_1})$, $\max_{1 \leq k \leq n} \left\{ \int_R \bar{F}_k(x) \cdot F_k(x) dx \right\}^2 = O(n^{\delta_2})$, $\sum_{1 \leq i \leq j \leq n} |\alpha_{ij}| = O(n^{\delta_3})$ for some $\delta_i > 0$, $i = 1, 2, 3$, then for all $\beta > \frac{1}{2} + \max\{\frac{\delta_1}{2}, \delta_2 + \delta_3\}$,

$$\frac{1}{n^\beta} \sum_{i=1}^n (X_i - EX_i) \rightarrow 0 \text{ Completely.}$$

Proof: By using Markov's inequality and P_3 , for all $\varepsilon > 0$, we can get

$$\begin{aligned} \sum_{n=1}^{\infty} P\left(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k (X_i - EX_i) \right| > n^\beta \varepsilon\right) &\leq \sum_{n=1}^{\infty} \frac{1}{n^{2\beta} \varepsilon^2} E\left(\sum_{i=1}^n (X_i - EX_i)\right)^2 \\ &\leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^{2\beta}} \sum_{i=1}^n \text{Var}(X_i) \\ &+ \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^{2\beta}} \sum_{i \neq j} |\alpha_{ij}| \max_{1 \leq k \leq n} \left\{ \int_R \bar{F}_k(x) \cdot F_k(x) dx \right\}^2 \end{aligned}$$

$$\leq \frac{1}{\varepsilon^2} \left\{ \sum_{n=1}^{\infty} \frac{1}{n^{2\beta-\delta_1}} + \sum_{n=1}^{\infty} \frac{1}{n^{2\beta-2\delta_2-2\delta_3}} \right\} < \infty.$$

This completes the proof.

References

- Amblard, C. and Girard, S.(2009) A new extension of bivariate FGM copulas. *Metrika*. 70, 1-17.
- Amini, M., Bozorgnia, A. (2000). Negatively dependent bounded random variable probability inequalities and the SLLN . *J.Appl.Math.Stoch.* 13:3, 261-267.
- Amini, M., Nili, H., Bozorgnia, A.(2012). Complete convergence of moving-average processes under negative dependent Sub-Gaussian assumptions. *Bull.Iran.Math.Soci.* 38(3), 843-852.
- Amini, M., Bozorgnia, A., Naderi, H. and Volodin, A. (2015) complete convergence of moving-average processes based on an identically distributed doubly infinite sequence of negatively superadditive-dependent random variables. *Sib.Adv.Math.* 25(1), 11-20.
- Baker, R.(2008)An order-statistics-based method for constructing multivariate distributions with fixed marginals. *J.Multiv.Anal.*99, 2312-2327.
- Cambanis, S.(1977) Some properties and generalizations of multivariate FGM distributions. *J.Multi.Anal.*7,551-559
- Cambanis, S.(1993) On EFGM random processes , in *Stochastic Inequalities IMS Lecture notes Monograph series*, Vol.22, Institut of Mathematical Statistics, Hayward, CA, 1993.
- Erdos P., (1949). On a theorem of Hsu and Robbins. *The Annals of Mathematical Statistics*, 20 (2), 286-291.
- Fisher, M. and Klein, I.(2007) Constructing generalized FGM copulas by means of certain univariate distributions. *Metrika*,65, 243-260.
- Gao, Sh., Zhang, Jin. Zhou, Tao.(2003). Law of large numbers for sample mean of random weighting estimate. *Info. sci.* 155,151-156.

- Hashorva, E. and Husler, J. (1999) Extreme values in FGM random sequences. *J. Multiv. Anal.* 68, 212-225.
- Hashorva, E. (2001), Asymptotic results for FGM random sequences. *Stat. Probab. Lett.* 54, 417-425.
- Hsu, P.L., Robbins, H. (1974). Complete convergence and the law of large numbers. *Proc. Nat. Acad.* 33, 25-31.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Stat. Assoc.* 58, 13-30.
- Johnson, N.L. and Kotz, S. (1977) On some generalized FGM distributions-II Regression, correlation and further generalizations. *Commun. Statist. Theor. Meth.* A6(6), 485-496.
- Liu, L. (2009). Precise large deviations for dependent random variables with heavy tails. *Stat. Probab. Lett.* 99, 1290-1298.
- Mikami, T. (1997). Large deviations and central limit theorems for EFGM processes. *Statist. Probab. Lett.* 35, 73-78.
- Mirhossini, S.M., Amini, M. and Dolati, A. (2015). On a general structure of the bivariate FGM type distributions. *Appl. Math.* 60(1), 91-108.
- Nadarajah, S. (2015). Expansions for bivariate copulas. *Statist. Probab. Lett.* 100, 77-84.
- Naderi, H., Amini, M. and Bozorgnia, A. (2015) Complete Convergence for Weighted Sums of Weakly Negative Dependent of Random Variables. *Filo Math.* Accepted.
- Nelsen, R. (2006). An introduction to copulas. Springer, New York.
- Ranjbar, V., Amini, M., Bozorgnia, A. (2009). Asymptotic behavior of convolution of dependent random variables with heavy-tail distribution. *Thai. J. Math.* 7, 21-34.
- Shen, A., Hu, S., Volodin, A., Wang, X. (2011). Some exponential inequalities for acceptable random variables and complete convergence. *J. Inequal. Appl.* 142.

- Sklar, A.(1959). Fonctions de répartition á n dimen-sions e leurs marges. Publications de l'Institut de Statistique de l'Univiversite de Paris.8:229-231.
- Sung, S.H. (2012). Complete convergence for weighted sums of negatively dependent random variables. Stat.Paper. 53:73-82.
- Wang, X., Hu, Sh., Yang, W. (2012). Complete convergence for arrays of rowwise negatively orthant dependent random variables. RACSAM. 106:235-245.
- Wang X.J., Deng X., Zheng L.L., Hu S.H., (2014). Complete convergence for arrays of rowwise negatively superadditive-dependent random variables and its applications. Statistics: 48(4),834-850.
- Wu, Q. (2012). A complete convergence theorem for weighted sums of arrays of rowwise negatively dependent random variables. J.Inequal.Appl. 50.
- Zheng, Z. (1987). Random weighting method. Acta.Math.Appli.Sinica. 10(2), 247-253.



A High Order Finite Difference Scheme for Stochastic Advection Diffusion Equations

M. Namjoo⁹, A. Mohebbian²

^{1,2}Departement of Mathematics, School of Mathematical Sciences, Vali-e-Asr
University of Rafsanjan, Rafsanjan, Iran.

Abstract: In this paper for the numerical approximation of stochastic advection diffusion equations, an explicit higher order finite difference scheme is constructed. In continuation the main properties of stochastic difference schemes, i.e. consistency, stability and convergency, are established for proposed stochastic difference scheme.

Keywords Stochastic partial differential equations, Consistency, Stability, Convergence.

Mathematics Subject Classification (2010): 34A12, 34D20.

1 Introduction

Physical phenomena of interest in science and technology are very often theoretically simulated by means of models which correspond to stochastic partial differential equations (SPDEs). As only a few, very simple SPDEs can be solved analytically, as a consequence, there is the need for the numerical schemes for approximating their solution. In recent years, some of numerical methods for solving SPDEs like finite difference and finite elements methods have been applied to the linear SPDEs, and the results of these approaches have been experimented numerically [Allen et al. \(1998\)](#). SPDEs have many applications in continuum physics [Bellmo et al. \(1989\)](#), mathematical biology and finance. In this paper an explicit finite difference scheme to approximate of stochastic advection diffusion equations is introduced and consistency, stability and convergency of this scheme is investigated.

⁹M. Namjoo : namjoo@vru.ac.ir

2 Finite Difference Approximation for Advection Diffusion Equations

Consider the following stochastic advection diffusion equation

$$u_t(x, t) + \nu u_x(x, t) = \gamma u_{xx}(x, t) + \sigma u(x, t) \dot{W}(t), \quad x \in [0, X], \quad t \in [0, T], \quad (2.1)$$

$$u(x, 0) = u_0(x), \quad u(0, t) = u(X, t) = 0,$$

where ν , γ and σ are positive random variables such that $\mathbb{E}(\nu^2)$, $\mathbb{E}(\gamma^2)$, $\mathbb{E}(\sigma^2) < \infty$, and $W(t)$ is a one-dimensional standard Wiener process such that the white noise $\dot{W}(t)$ is a Gaussian distribution with zero mean [Kloeden et al. \(1998\)](#). Numerically, finite difference methods have vast applications in approximating the solution of SPDEs. These schemes discretize continuous space and time into an evenly distributed grid system, and the values of the state variables are evaluated at each node of the grid. Considering a uniform space grid Δx and time grid Δt in the time-space lattice, we can estimate the solution of the equation at the points of this lattice. The value of the approximate solution at the point $(k\Delta x, n\Delta t)$ will be denoted by u_k^n where n, k are integers. In the explicit method, the time and space derivatives in the SPDE are approximated by finite difference replacements in the following form [Thomas \(1998\)](#):

$$u_x(k\Delta x, n\Delta t) \approx \frac{u_{k+1}^n - u_k^n}{\Delta x}, \quad u_t(k\Delta x, n\Delta t) \approx \frac{u_k^{n+1} - u_k^n}{\Delta t},$$

$$u_{xx}(x, t) \approx \frac{1}{\Delta x^2} \left(-\frac{1}{12}u(x - 2\Delta x, t) + \frac{4}{3}u(x - \Delta x, t) - \frac{5}{2}u(x, t) + \frac{4}{3}u(x + \Delta x, t) - \frac{1}{12}u(x + 2\Delta x, t) \right).$$

Therefore, the scheme approximates the stochastic advection diffusion equation is given by

$$u_k^{n+1} = \left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right) u_k^n + \left(\frac{4}{3}\gamma\rho - \nu\lambda \right) u_{k+1}^n + \gamma\rho \left(-\frac{1}{12}u_{k-2}^n + \frac{4}{3}u_{k-1}^n - \frac{1}{12}u_{k+2}^n \right) + \sigma u_k^n \Delta W_n, \quad (2.2)$$

where $\lambda = \frac{\Delta t}{\Delta x}$, $\rho = \frac{\Delta t}{\Delta x^2}$ and $\Delta W_n = W((n+1)\Delta t) - W(n\Delta t)$ is a Gaussian distribution with mean 0 and variance Δt , i.e. $\Delta W_n \sim N(0, \Delta t)$.

Remark 2.1. For the proposed scheme, we assume that the random variables ν , γ and σ are independent of the Wiener process and the states u_k^n .

Consider a SPDE in the following form $Lv = G$, where L denotes the differential operator and $G \in L^2(\mathbb{R})$ is an inhomogeneity. Assuming u_k^n is the solution that is approximated by a stochastic finite difference scheme denoted by L_k^n , and applying the stochastic scheme to the SPDE, we have $L_k^n u_k^n = G_k^n$, where G_k^n is the approximation of the inhomogeneity. For consistency, stability and convergence, we will need a norm. Hence for a sequence $u = \{\dots, u_{-1}, u_0, u_1, \dots\}$, the sup-norm is defined as $\|u\|_\infty = \sqrt{\sup_k |u_k|^2}$. We refer to the paper [Roth \(2008\)](#), for the following definitions of stochastic difference scheme.

Definition 2.2. A stochastic difference scheme $L_k^n u_k^n = G_k^n$ is pointwise consistent with the stochastic partial differential equation $Lv = G$ at point (x, t) , if for any continuously differentiable function $\Phi = \Phi(x, t)$, in mean square

$$\mathbb{E} \| (L\Phi - G)|_k^n - [L_k^n \Phi(k\Delta x, n\Delta t) - G_k^n] \|^2 \rightarrow 0,$$

as $\Delta x \rightarrow 0$, $\Delta t \rightarrow 0$, and $(k\Delta x, (n+1)\Delta t)$ converges to (x, t) .

Definition 2.3. A stochastic difference scheme is said to be stable with respect to a norm in mean square if there exist some positive constants $\overline{\Delta x_0}$ and $\overline{\Delta t_0}$ and non-negative constants K and β such that

$$\mathbb{E} \|u^{n+1}\|^2 \leq K e^{\beta t} \mathbb{E} \|u^0\|^2,$$

for all $0 \leq t = (n+1)\Delta t$, $0 \leq \Delta x \leq \overline{\Delta x_0}$, and $0 \leq \Delta t \leq \overline{\Delta t_0}$, where

$$u^{n+1} = (\dots, u_{k-2}^{n+1}, u_{k-1}^{n+1}, u_k^{n+1}, u_{k+1}^{n+1}, u_{k+2}^{n+1}, \dots)^T.$$

Definition 2.4. A stochastic difference schemes $L_k^n u_k^n = G_k^n$ approximating the stochastic partial differential equation $Lv = G$ is convergent in mean square at time t if as $(n+1)\Delta t$ converges to t , $\mathbb{E} \|u^{n+1} - v^{n+1}\|^2 \rightarrow 0$, for $(n+1)\Delta t = t$, $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$.

3 Consistency, Stability and Convergence of Stochastic Scheme

Theorem 3.1. The stochastic difference scheme (2.2) is consistent in mean square in the sense of Definition 2.2.

Proof. Let $\Phi(x, t)$ be a smooth function. Then we have:

$$L(\Phi)|_k^n = \Phi(k\Delta x, (n+1)\Delta t) - \Phi(k\Delta x, n\Delta t) + \nu \int_{n\Delta t}^{(n+1)\Delta t} \Phi_x(k\Delta x, s) ds$$

$$-\gamma \int_{n\Delta t}^{(n+1)\Delta t} \Phi_{xx}(k\Delta x, s) ds - \sigma \int_{n\Delta t}^{(n+1)\Delta t} \Phi(k\Delta x, s) dW(s),$$

and

$$\begin{aligned} L_k^n \Phi &= \Phi(k\Delta x, (n+1)\Delta t) - \Phi(k\Delta x, n\Delta t) + \nu \Delta t \left(\frac{\Phi((k+1)\Delta x, n\Delta t) - \Phi(k\Delta x, n\Delta t)}{\Delta x} \right) \\ &\quad - \gamma \frac{\Delta t}{\Delta x^2} \left(-\frac{1}{12} \Phi((k-2)\Delta x, n\Delta t) + \frac{4}{3} \Phi((k-1)\Delta x, n\Delta t) - \frac{5}{2} \Phi(k\Delta x, n\Delta t) \right. \\ &\quad \left. + \frac{4}{3} \Phi((k+1)\Delta x, n\Delta t) - \frac{1}{12} \Phi((k+2)\Delta x, n\Delta t) \right) - \sigma \Phi(k\Delta x, n\Delta t) (W((n+1)\Delta t) - W(n\Delta t)). \end{aligned}$$

Therefore, in mean square, we obtain:

$$\begin{aligned} &\mathbb{E}|L(\Phi)|_k^n - L_k^n \Phi|^2 \\ &= \mathbb{E} \left| \nu \int_{n\Delta t}^{(n+1)\Delta t} \left(\Phi_x(k\Delta x, s) - \frac{\Phi((k+1)\Delta x, n\Delta t) - \Phi(k\Delta x, n\Delta t)}{\Delta x} \right) ds \right. \\ &\quad - \gamma \int_{n\Delta t}^{(n+1)\Delta t} \left(\Phi_{xx}(k\Delta x, s) - \frac{1}{\Delta x^2} \left[-\frac{1}{12} \Phi((k-2)\Delta x, n\Delta t) + \frac{4}{3} \Phi((k-1)\Delta x, n\Delta t) \right. \right. \\ &\quad \left. \left. - \frac{5}{2} \Phi(k\Delta x, n\Delta t) + \frac{4}{3} \Phi((k+1)\Delta x, n\Delta t) - \frac{1}{12} \Phi((k+2)\Delta x, n\Delta t) \right] \right) ds \\ &\quad \left. - \sigma \int_{n\Delta t}^{(n+1)\Delta t} (\Phi(k\Delta x, s) - \Phi(k\Delta x, n\Delta t)) dW(s) \right|^2 \\ &\leq \mathbb{E}(\nu^2) \mathbb{E} \left| \int_{n\Delta t}^{(n+1)\Delta t} \left(\Phi_x(k\Delta x, s) - \frac{\Phi((k+1)\Delta x, n\Delta t) - \Phi(k\Delta x, n\Delta t)}{\Delta x} \right) ds \right|^2 \\ &\quad + 4\mathbb{E}(\gamma^2) \mathbb{E} \left| \int_{n\Delta t}^{(n+1)\Delta t} \left(\Phi_{xx}(k\Delta x, s) - \frac{1}{\Delta x^2} \left[-\frac{1}{12} \Phi((k-2)\Delta x, n\Delta t) + \frac{4}{3} \Phi((k-1)\Delta x, n\Delta t) \right. \right. \right. \\ &\quad \left. \left. - \frac{5}{2} \Phi(k\Delta x, n\Delta t) + \frac{4}{3} \Phi((k+1)\Delta x, n\Delta t) - \frac{1}{12} \Phi((k+2)\Delta x, n\Delta t) \right] \right) ds \right|^2 \\ &\quad + 4\mathbb{E}(\sigma^2) \int_{n\Delta t}^{(n+1)\Delta t} \mathbb{E}|\Phi(k\Delta x, s) - \Phi(k\Delta x, n\Delta t)|^2 ds \end{aligned}$$

Since $\Phi(x, t)$ is a deterministic function, hence $\mathbb{E}|L(\Phi)|_k^n - L_k^n \Phi|^2 \rightarrow 0$, as $n, k \rightarrow \infty$. \square

Theorem 3.2. *The stochastic difference scheme (2.2) with $t = (n+1)\Delta t$ and $\frac{3}{4}\nu\lambda \leq \gamma\rho \leq \frac{2}{5}(1+\nu\lambda)$, (note that $\nu\lambda \leq \frac{8}{7}$), is stable with respect to $\|\cdot\|_\infty = \sqrt{\sup_k |\cdot|^2}$.*

Proof. Applying $\mathbb{E}|\cdot|^2$ in (2.2) and using the independence of the Wiener process increments we will have:

$$\begin{aligned} \mathbb{E}|u_k^{n+1}|^2 &= \mathbb{E} \left| \left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right) u_k^n + \left(\frac{4}{3}\gamma\rho - \nu\lambda \right) u_{k+1}^n + \gamma\rho \left(-\frac{1}{12}u_{k-2}^n + \frac{4}{3}u_{k-1}^n - \frac{1}{12}u_{k+2}^n \right) \right|^2 \\ &\quad + \mathbb{E}(\sigma^2)\Delta t \mathbb{E}|u_k^n|^2. \end{aligned}$$

By using $\frac{3}{4}\nu\lambda \leq \gamma\rho \leq \frac{2}{5}(1 + \nu\lambda)$, we get

$$\begin{aligned} \mathbb{E}|u_k^{n+1}|^2 &\leq \mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right)^2 \right) \mathbb{E}|u_k^n|^2 + \mathbb{E} \left(\left(\frac{4}{3}\gamma\rho - \nu\lambda \right)^2 \right) \mathbb{E}|u_{k+1}^n|^2 \\ &\quad + \mathbb{E}((\gamma\rho)^2) \mathbb{E} \left| -\frac{1}{12}u_{k-2}^n + \frac{4}{3}u_{k-1}^n - \frac{1}{12}u_{k+2}^n \right|^2 + 2\mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right) \left(\frac{4}{3}\gamma\rho - \nu\lambda \right) \right) \mathbb{E}|u_k^n u_{k+1}^n| \\ &\quad + 2\mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right) \gamma\rho \right) \mathbb{E} \left| u_k^n \left(-\frac{1}{12}u_{k-2}^n + \frac{4}{3}u_{k-1}^n - \frac{1}{12}u_{k+2}^n \right) \right| \\ &\quad + 2\mathbb{E} \left(\left(\frac{4}{3}\gamma\rho - \nu\lambda \right) \gamma\rho \right) \mathbb{E} \left| u_{k+1}^n \left(-\frac{1}{12}u_{k-2}^n + \frac{4}{3}u_{k-1}^n - \frac{1}{12}u_{k+2}^n \right) \right| + \mathbb{E}(\sigma^2)\Delta t \mathbb{E}|u_k^n|^2. \end{aligned}$$

Hence, using the inequalities

$$\mathbb{E}|X + Y + Z| \leq \mathbb{E}|X| + \mathbb{E}|Y| + \mathbb{E}|Z|, \quad \mathbb{E}|X + Y + Z|^2 \leq 4(\mathbb{E}|X|^2 + \mathbb{E}|Y|^2 + \mathbb{E}|Z|^2),$$

we will have:

$$\begin{aligned} \mathbb{E}|u_k^{n+1}|^2 &\leq \mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right)^2 \right) \mathbb{E}|u_k^n|^2 + \mathbb{E} \left(\left(\frac{4}{3}\gamma\rho - \nu\lambda \right)^2 \right) \mathbb{E}|u_{k+1}^n|^2 \\ &\quad + \frac{\mathbb{E}((\gamma\rho)^2)}{144} (4\mathbb{E}|u_{k-2}^n|^2 + 1024\mathbb{E}|u_{k-1}^n|^2 + 4\mathbb{E}|u_{k+2}^n|^2) \\ &\quad + 2\mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right) \left(\frac{4}{3}\gamma\rho - \nu\lambda \right) \right) \mathbb{E}|u_k^n u_{k+1}^n| \\ &\quad + \frac{1}{6}\mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right) \gamma\rho \right) (\mathbb{E}|u_k^n u_{k-2}^n| + 16\mathbb{E}|u_k^n u_{k-1}^n| + \mathbb{E}|u_k^n u_{k+2}^n|) \\ &\quad + \frac{1}{6}\mathbb{E} \left(\left(\frac{4}{3}\gamma\rho - \nu\lambda \right) \gamma\rho \right) (\mathbb{E}|u_{k+1}^n u_{k-2}^n| + 16\mathbb{E}|u_{k+1}^n u_{k-1}^n| + \mathbb{E}|u_{k+1}^n u_{k+2}^n|) + \mathbb{E}(\sigma^2)\Delta t \mathbb{E}|u_k^n|^2 \\ &\leq \left\{ \mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right)^2 \right) + \mathbb{E} \left(\left(\frac{4}{3}\gamma\rho - \nu\lambda \right)^2 \right) + \frac{1032}{144}\mathbb{E}((\gamma\rho)^2) \right\} \end{aligned}$$

$$\begin{aligned}
 &+2\mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right) \left(\frac{4}{3}\gamma\rho - \nu\lambda \right) \right) + 3\mathbb{E} \left(\left(1 + \nu\lambda - \frac{5}{2}\gamma\rho \right) \gamma\rho \right) \\
 &\quad + 3\mathbb{E} \left(\left(\frac{4}{3}\gamma\rho - \nu\lambda \right) \gamma\rho \right) + \mathbb{E}(\sigma^2)\Delta t \Big\} \sup_k \mathbb{E}|u_k^n|^2 \\
 &= \left\{ 1 + \frac{2}{3}\mathbb{E}(\gamma\rho) + \frac{181}{36}\mathbb{E}((\gamma\rho)^2) + \mathbb{E}(\sigma^2)\Delta t \right\} \sup_k \mathbb{E}|u_k^n|^2.
 \end{aligned}$$

It is enough to select δ such that $\frac{2}{3}\mathbb{E}(\gamma\rho) + \frac{181}{36}\mathbb{E}((\gamma\rho)^2) + \mathbb{E}(\sigma^2)\Delta t \leq \delta^2\Delta t$ holds, for all k . Therefore

$$\sup_k \mathbb{E}|u_k^{n+1}|^2 \leq (1 + \delta^2\Delta t) \sup_k \mathbb{E}|u_k^n|^2 \leq \dots \leq (1 + \delta^2\Delta t)^{n+1} \sup_k \mathbb{E}|u_k^0|^2,$$

and by substituting Δt with $\frac{t}{n+1}$,

$$\mathbb{E}\|u^{n+1}\|_\infty^2 \leq \left(1 + \frac{\delta^2 t}{n+1} \right)^{n+1} \mathbb{E}\|u^0\|_\infty^2 \leq e^{\delta^2 t} \mathbb{E}\|u^0\|_\infty^2.$$

Hence, the stochastic difference scheme is conditionally stable with $K = 1$ and $\beta = \delta^2$. □

Theorem 3.3. *The stochastic difference scheme (2.2) is convergent with respect to $\|\cdot\|_\infty$ -norm, for $\frac{3}{4}\nu\lambda \leq \gamma\rho \leq \frac{2}{5}(1 + \nu\lambda)$.*

Proof. The stochastic finite difference scheme is given by

$$\begin{aligned}
 u_k^{n+1} &= u_k^n - \nu\Delta t \frac{u_{k+1}^n - u_k^n}{\Delta x} + \gamma \frac{\Delta t}{\Delta x^2} \left(-\frac{1}{12}u_{k-2}^n + \frac{4}{3}u_{k-1}^n - \frac{5}{2}u_k^n + \frac{4}{3}u_{k+1}^n - \frac{1}{12}u_{k+2}^n \right) \\
 &\quad + \sigma u_k^n (W((n+1)\Delta t) - W(n\Delta t)).
 \end{aligned}$$

The solution v_k^{n+1} can be represented by the Taylor expansion $v_{xx}(x, s)$ with respect to the space variable as

$$\begin{aligned}
 v_k^{n+1} &= v_k^n - \nu \int_{n\Delta t}^{(n+1)\Delta t} v_x(x, s)|_{x=x_k} ds + \gamma \int_{n\Delta t}^{(n+1)\Delta t} v_{xx}(x_k, s) ds \\
 &\quad + \sigma \int_{n\Delta t}^{(n+1)\Delta t} v(x, s)|_{x=x_k} dW(s) \\
 &= v_k^n - \nu \int_{n\Delta t}^{(n+1)\Delta t} \left(\frac{v_{k+1}^n - v_k^n}{\Delta x} - \frac{\Delta x}{2} v_{xx}((k + \alpha)\Delta x, s) \right) ds \\
 &\quad + \gamma \int_{n\Delta t}^{(n+1)\Delta t} \left(\frac{1}{\Delta x^2} \left(-\frac{1}{12}v_{k-2}^n + \frac{4}{3}v_{k-1}^n - \frac{5}{2}v_k^n + \frac{4}{3}v_{k+1}^n - \frac{1}{12}v_{k+2}^n \right) \right. \\
 &\quad \left. + \frac{\Delta x^4}{135} \left(v^{(6)}((k + \theta)\Delta x, s) + v^{(6)}((k + \mu)\Delta x, s) \right) - \frac{\Delta x^4}{540} \left(v^{(6)}((k + \theta')\Delta x, s) + v^{(6)}((k + \mu')\Delta x, s) \right) \right) ds
 \end{aligned}$$

$$+\sigma \int_{n\Delta t}^{(n+1)\Delta t} v(x, s)|_{x=x_k} dW(s),$$

where $\alpha, \theta, \mu, \theta', \mu' \in (0, 1)$. With $z_k^n = v_k^n - u_k^n$ we get

$$\begin{aligned} z_k^{n+1} &= \left(1 + \nu\lambda - \frac{5}{2}\gamma\rho\right) z_k^n + \left(\frac{4}{3}\gamma\rho - \nu\lambda\right) z_{k+1}^n + \gamma\rho \left(-\frac{1}{12}z_{k-2}^n + \frac{4}{3}z_{k-1}^n - \frac{1}{12}z_{k+2}^n\right) \\ &+ \nu \frac{\Delta x}{2} \int_{n\Delta t}^{(n+1)\Delta t} v_{xx}((k + \alpha)\Delta x, s) ds - \gamma \int_{n\Delta t}^{(n+1)\Delta t} \left(\frac{\Delta x^4}{135} \left(v^{(6)}((k + \theta)\Delta x, s) + v^{(6)}((k + \mu)\Delta x, s)\right) \right. \\ &\left. - \frac{\Delta x^4}{540} \left(v^{(6)}((k + \theta')\Delta x, s) + v^{(6)}((k + \mu')\Delta x, s)\right)\right) ds + \sigma \int_{n\Delta t}^{(n+1)\Delta t} (v(x, s)|_{x=x_k} - u_k^n) dW(s). \end{aligned}$$

Applying $\mathbb{E}|\cdot|^2$ to this equality, we obtain

$$\begin{aligned} \mathbb{E}|z_k^{n+1}|^2 &\leq 4\mathbb{E} \left| \left(1 + \nu\lambda - \frac{5}{2}\gamma\rho\right) z_k^n + \left(\frac{4}{3}\gamma\rho - \nu\lambda\right) z_{k+1}^n + \gamma\rho \left(-\frac{1}{12}z_{k-2}^n + \frac{4}{3}z_{k-1}^n - \frac{1}{12}z_{k+2}^n\right) \right|^2 \\ &+ 4\mathbb{E} \left| \nu \frac{\Delta x}{2} \int_{n\Delta t}^{(n+1)\Delta t} v_{xx}((k + \alpha)\Delta x, s) ds - \gamma \int_{n\Delta t}^{(n+1)\Delta t} \left(\frac{\Delta x^4}{135} \left(v^{(6)}((k + \theta)\Delta x, s) + v^{(6)}((k + \mu)\Delta x, s)\right) \right. \right. \\ &\left. \left. - \frac{\Delta x^4}{540} \left(v^{(6)}((k + \theta')\Delta x, s) + v^{(6)}((k + \mu')\Delta x, s)\right)\right) ds \right|^2 + 2\mathbb{E} \left| \sigma \int_{n\Delta t}^{(n+1)\Delta t} (v(x, s)|_{x=x_k} - u_k^n) dW(s) \right|^2 \\ &\leq 4\mathbb{E} \left| \left(1 + \nu\lambda - \frac{5}{2}\gamma\rho\right) z_k^n + \left(\frac{4}{3}\gamma\rho - \nu\lambda\right) z_{k+1}^n + \gamma\rho \left(-\frac{1}{12}z_{k-2}^n + \frac{4}{3}z_{k-1}^n - \frac{1}{12}z_{k+2}^n\right) \right|^2 \\ &+ 4\mathbb{E} \left| \nu \frac{\Delta x}{2} \int_{n\Delta t}^{(n+1)\Delta t} v_{xx}((k + \alpha)\Delta x, s) ds - \gamma \int_{n\Delta t}^{(n+1)\Delta t} \left(\frac{\Delta x^4}{135} \left(v^{(6)}((k + \theta)\Delta x, s) + v^{(6)}((k + \mu)\Delta x, s)\right) \right. \right. \\ &\left. \left. - \frac{\Delta x^4}{540} \left(v^{(6)}((k + \theta')\Delta x, s) + v^{(6)}((k + \mu')\Delta x, s)\right)\right) ds \right|^2 + 2\mathbb{E}(\sigma^2) \int_{n\Delta t}^{(n+1)\Delta t} \mathbb{E}|v(x, s)|_{x=x_k} - v_k^n + v_k^n - u_k^n|^2 ds. \end{aligned}$$

Assuming $\frac{3}{4}\nu\lambda \leq \gamma\rho \leq \frac{2}{5}(1 + \nu\lambda)$, introducing the notation $\psi_{1k} = v_{xx}((k + \alpha)\Delta x, s) < \infty$, $\psi_{2k} = v^{(6)}((k + \theta)\Delta x, s) < \infty$, $\psi_{3k} = v^{(6)}((k + \mu)\Delta x, s) < \infty$, $\psi_{4k} = v^{(6)}((k + \theta')\Delta x, s) < \infty$, $\psi_{5k} = v^{(6)}((k + \mu')\Delta x, s) < \infty$ and also

$$\begin{aligned} \int_{n\Delta t}^{(n+1)\Delta t} \mathbb{E}|v(x, s)|_{x=x_k} - v_k^n|^2 ds &= \mathbb{E} \int_{n\Delta t}^{(n+1)\Delta t} |v(x, s)|_{x=x_k} - v_k^n|^2 ds \\ &\leq \sup_{s \in [n\Delta t, (n+1)\Delta t]} |v(x, s)|_{x=x_k} - v(k\Delta x, n\Delta t)|^2 \Delta t \leq \psi' \Delta t, \end{aligned}$$

we get

$$\begin{aligned} \mathbb{E}|z_k^{n+1}|^2 &\leq 4 \left(1 + \frac{2}{3}\mathbb{E}(\gamma\rho) + \frac{181}{36}\mathbb{E}((\gamma\rho)^2) \right) \sup_k \mathbb{E}|z_k^n|^2 \\ &+ 4 \sup_k \mathbb{E} \left| \int_{n\Delta t}^{(n+1)\Delta t} \left(\nu \frac{\Delta x}{2} \psi_{1k} - \gamma \frac{\Delta x^4}{135} (\psi_{2k} + \psi_{3k}) + \gamma \frac{\Delta x^4}{540} (\psi_{4k} + \psi_{5k}) \right) ds \right|^2 \\ &+ 4\mathbb{E}(\sigma^2) \sup_k \int_{n\Delta t}^{(n+1)\Delta t} \mathbb{E}|v(x, s)|_{x=x_k} - v_k^n|^2 ds + 4\mathbb{E}(\sigma^2) \underbrace{\sup_k \int_{n\Delta t}^{(n+1)\Delta t} \mathbb{E}|v_k^n - u_k^n|^2 ds}_{\mathbb{E}|z_k^n|^2 \Delta t} \\ &\leq 4 \left(1 + \frac{2}{3}\mathbb{E}(\gamma\rho) + \frac{181}{36}\mathbb{E}((\gamma\rho)^2) + \mathbb{E}(\sigma^2)\Delta t \right) \sup_k \mathbb{E}|z_k^n|^2 + 4 \sup_k \mathbb{E}|\Psi_1 \Delta t|^2 + \psi' \Delta t \\ &\leq 4 \left(1 + \frac{2}{3}\mathbb{E}(\gamma\rho) + \frac{181}{36}\mathbb{E}((\gamma\rho)^2) + \mathbb{E}(\sigma^2)\Delta t \right) \sup_k \mathbb{E}|z_k^n|^2 + \Psi \Delta t. \end{aligned}$$

With selecting δ such that $\frac{2}{3}\mathbb{E}(\gamma\rho) + \frac{181}{36}\mathbb{E}((\gamma\rho)^2) + \mathbb{E}(\sigma^2)\Delta t \leq \delta^2 \Delta t$, we have:

$$\mathbb{E}|z_k^{n+1}|^2 \leq 4(1 + \delta^2 \Delta t) \sup_k \mathbb{E}|z_k^n|^2 + \Psi \Delta t,$$

and

$$\sup_k \mathbb{E}|z_k^{n+1}|^2 \leq 4(1 + \delta^2 \Delta t) \sup_k \mathbb{E}|z_k^n|^2 + \Psi \Delta t.$$

It follows that

$$\begin{aligned} \mathbb{E}\|z^{n+1}\|_\infty^2 &\leq 4(1 + \delta^2 \Delta t) \mathbb{E}\|z^n\|_\infty^2 + \Psi \Delta t \leq \left(1 + \delta^2 \frac{t}{n+1} \right)^{n+1} \sum_{j=1}^n (4\Psi \Delta t)^j + \Psi \Delta t \\ &\leq e^{\delta^2 t} \sum_{j=1}^n (4\Psi \Delta t)^j + \Psi \Delta t. \end{aligned}$$

When the time step Δt tends to zero, we have

$$\begin{aligned} \mathbb{E}\|z^{n+1}\|_\infty^2 &\leq (n-1)e^{\delta^2 t} (4\Psi \Delta t)^2 + 4e^{\delta^2 t} \Psi \Delta t + \Psi \Delta t \leq te^{\delta^2 t} (4\Psi)^2 \Delta t + 4e^{\delta^2 t} \Psi \Delta t + \Psi \Delta t \\ &= (te^{\delta^2 t} (4\Psi)^2 + 4e^{\delta^2 t} \Psi + \Psi) \Delta t, \end{aligned}$$

and consequently $\mathbb{E}\|z^{n+1}\|_\infty^2 \rightarrow 0$. □

Conclusion

In this paper the stochastic five points finite difference method is used to obtain an approximation solution of the stochastic advection diffusion equation. Consistency of the stochastic finite difference scheme is established. Sufficient conditions for the stability of the proposed stochastic difference scheme are given. Finally the convergence of the solution scheme to the exact one is proved.

References

- Allen, E. J., Novosel, S. J. and Zhang, Z. (1998). Finite element and difference approximation of some linear stochastic partial differential equations, *Stochastics Rep.*, **64**, 117–142.
- Bellmo, N. and Flandoli, F. (1989). Stochastic partial differential equations in continuum physics on the foundations of the stochastic interpolation method for Itô type equations, *Math. Comput. Simulation*, **31**, 3–17.
- Kloeden, P. E. and Platen, E. (1995). *Numerical solution of stochastic differential equations*, Springer, Berlin.
- Roth, C. (2002). Difference methods for stochastic partial differential equations, *Z. Zngew. Math. Mech.*, **82**, 821–830.
- Thomas, J. W. (1998). *Numerical partial differential equations: finite difference methods*, Springer.



The discrete beta generalized exponential distribution

Vahid Nekoukhou¹

Department of Statistics, University of Isfahan, Khansar Unit, Isfahan, Iran

Abstract: In this paper, a discrete analogue of the beta generalized exponential distribution is studied. This new distribution contains some previously known discrete distributions as well as two new models. The hazard rate function of the new model can be increasing, decreasing, bathtub-shaped and upside-down bathtub. Some distributional and moment properties of the new distribution are discussed and, finally, the model with a real data set is examined.

Keywords Beta-G distributions, Beta generalized exponential distribution, Discrete generalized exponential distribution, Hazard rate function.

Mathematics Subject Classification (2010): 11J71.

1 Introduction

Eugene et al. (2002) introduced a general class of distributions generated from the logit of a beta random variable with cumulative distribution function (cdf)

$$F(x; a, b, \theta) = I_{G(x)}(a, b) = \frac{1}{B(a, b)} \int_0^{G(x)} t^{a-1} (1-t)^{b-1} dt, \quad x \in \mathbb{R}, \quad (1.1)$$

where $a > 0$ and $b < \infty$ are two parameters whose role is to control skewness and the tail weight, θ is the parameter vector of the absolutely continuous cdf G , $I_y(a, b) = \frac{B_y(a, b)}{B(a, b)}$ is the incomplete beta function ratio and $B_y(a, b) = \int_0^y t^{a-1} (1-t)^{b-1} dt$ denotes the incomplete beta function. In fact, if U is a beta random variable with parameters a and b , then the cdf of the random variable $X = G^{-1}(U)$ coincides with cdf (1.1). The distribution function (1.1) is called a *beta-G* (BG) distribution and is denoted by $BG(a, b, \theta)$.

¹Speaker's Name : v.nekoukhou@gmail.com

In recent years, the class of BG distributions has received much attention in the literature. For instance, Cordeiro et al. (2013) considered the beta exponentiated Weibull distribution and discussed some of its basic distributional properties. Bidram et al. (2013) introduced the beta Weibull-geometric distribution and investigated some useful general results regarding the order statistics of the family of BG distributions. Beta generalized Pareto distribution of Mahmoudi (2011) is another research in this regard.

In this paper, we will introduce a discrete analog of the beta generalized exponential (BGE) distribution of Barreto-Souza et al. (2010). This new discrete distribution contains several special sub-models. Two of such sub-models are new distributions, which are introduced here. One of the new sub-models is a discrete analog of the beta exponential (BE) distribution of Nadarajah and Kotz (2006) and another is a discrete counterpart of the double generalized exponential distribution given, as a special sub-model of BGE, by Barreto-Souza et al. (2010).

2 The discrete beta generalized exponential distribution

Barreto-Souza et al. (2010) introduced the BGE distribution by inserting the cdf of the generalized exponential (GE) distribution of Gupta and Kundu (1999),

$$G(x) = (1 - e^{-\lambda x})^\alpha, \quad x > 0; \alpha > 0, \lambda > 0, \quad (2.1)$$

into Eq. (1.1). The pdf of the BGE distribution is given by

$$f(x) = \frac{\alpha\lambda}{B(a,b)} e^{-\lambda x} (1 - e^{-\lambda x})^{\alpha a - 1} \{1 - (1 - e^{-\lambda x})^\alpha\}^{b-1}, \quad x > 0. \quad (2.2)$$

The BGE distribution generalizes some well-known distributions in the literature. The GE distribution is a special case for the choice $a = b = 1$. If, in addition, $\alpha = 1$, the exponential distribution with parameter λ is obtained. The BE distribution is obtained by choosing $\alpha = 1$. The double generalized exponential distribution corresponds to $a = 1$.

The BGE distribution is much more flexible than the GE and BE distributions in view of the fact that its hazard rate function can be bathtub-shaped, monotonically increasing or decreasing, and upside-down bathtub depending on the values of the parameters; see Barreto-Souza et al. (2010).

Here, we construct a discrete analogue of the BGE distribution using equation

$$f_{(a,b,\theta)}(y) = P(Y = y) = I_{G(y+1)}(a, b) - I_{G(y)}(a, b), \quad (2.3)$$

as follows:

$$\begin{aligned} f_{(a,b,\theta)}(y) = p_y &= \sum_{i=0}^{\infty} \omega_i(a, b) \{ (1 - e^{-\lambda(y+1)})^{\alpha(a+i)} - (1 - e^{-\lambda y})^{\alpha(a+i)} \} \\ &= \sum_{i=0}^{\infty} \omega_i(a, b) \{ (1 - p^{y+1})^{\alpha(a+i)} - (1 - p^y)^{\alpha(a+i)} \} \end{aligned} \quad (2.4)$$

$$= \sum_{i=0}^{\infty} \omega_i(a, b) \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\alpha(a+i)}{j} p^{jy} (1 - p^j), \quad (2.5)$$

where $0 < p = e^{-\lambda} < 1$ and $y \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$. The pmf given in (2.3) or (2.4) defines the discrete beta generalized exponential distribution and we denote it by $DBGE(a, b, \alpha, p)$. The pmf p_y is in fact a linear combination of the probability mass functions of the discrete generalized exponential distributions of a second type, $DGE_2(\alpha(a+i), p)$, introduced by Nekoukhou et al. (2013a). A $DGE_2(\alpha, p)$ distribution is indeed a discrete counterpart of the two-parameter GE distribution, given by (2.1), whose pmf is of the form

$$p_y = (1 - p^{y+1})^{\alpha} - (1 - p^y)^{\alpha} \quad (2.6)$$

$$= \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\alpha}{j} p^{jy} (1 - p^j), \quad y \in \mathbb{N}_0, \quad (2.7)$$

where $\binom{\alpha}{j} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)j!}$. For integer $\alpha > 0$, the sum in Eq. (2.6) stops at α . The DGE_2 is indeed an exponentiated geometric distribution in view of the fact that the cdf of a random variable Y following the $DGE_2(\alpha, p)$ distribution is given by

$$P(Y \leq y) = \begin{cases} 0 & y < 0 \\ (1 - p^{[y]+1})^{\alpha} & y \geq 0. \end{cases} \quad (2.8)$$

It is interesting to note that the BGE distribution is also a linear combination of the GE density functions. In addition, the DBGE distribution is an infinite linear combination of DGE_2 distributions. Hence, we can obtain some basic properties of the DBGE distribution from the corresponding properties of DGE_2 distributions.

The cdf of a $DBGE(a, b, \alpha, p)$ distribution is given by

$$F_{(a,b,\alpha,p)}(y) = P(Y \leq y) = \begin{cases} 0 & y < 0 \\ \sum_{i=0}^{\infty} \omega_i(a, b)(1 - p^{[y]+1})^{\alpha(a+i)} & y \geq 0, \end{cases} \quad (2.9)$$

which is an infinite linear combination of the cdf of $DGE_2(\alpha(a + i), p)$ distributions. The survival and hazard rate functions of a random variable Y following a $DBGE(a, b, \alpha, p)$ distribution are given, respectively, by

$$\bar{F}_{(a,b,\alpha,p)}(y) = P(Y > y) = 1 - \sum_{i=0}^{\infty} \omega_i(a, b)(1 - p^{[y]+1})^{\alpha(a+i)}, \quad y \geq 0 \quad (2.10)$$

and

$$h_{(a,b,\alpha,p)}(y) = \frac{\sum_{i=0}^{\infty} \omega_i(a, b)\{(1 - p^{y+1})^{\alpha(a+i)} - (1 - p^y)^{\alpha(a+i)}\}}{1 - \sum_{i=0}^{\infty} \omega_i(a, b)(1 - p^{[y]+1})^{\alpha(a+i)}}, \quad y \in \mathbb{N}_0. \quad (2.11)$$

As mentioned before, for integer $b > 0$, $\sum_{i=0}^{\infty}$ should be replaced by $\sum_{i=0}^{b-1}$. Figure 1 illustrates the hazard rate function behavior of the DBGE distribution for different values of its parameters. As we see from Figure 1, a characteristic of a DBGE distribution is that its hazard rate function can be constant, decreasing, increasing, bathtub-shaped, and upside-down bathtub depending on its parameters values. Hence, DBGE distributions are more flexible than other discrete distributions such as the geometric or DGE_2 distributions, whose hazard rate functions are constant and monotone, respectively.

The mgf of a $DBGE(a, b, \alpha, p)$ distribution can be obtained easily by that of $DGE_2(\alpha(a + i), p)$ distributions. In fact, we have

$$M_Y(t) = \sum_{i=0}^{\infty} \omega_i(a, b) \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\alpha(a+i)}{j} \frac{1 - p^j}{1 - p^j e^t}, \quad t < -\log p. \quad (2.12)$$

The factorial moments and moments of a $DBGE(a, b, \alpha, p)$ distribution for $n \in \mathbb{N} = \{1, 2, \dots\}$ are obtained, respectively, as

$$E\{Y(Y - 1)\dots(Y - n + 1)\} = \sum_{i=0}^{\infty} \omega_i(a, b) \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\alpha(a+i)}{j} n! \left(\frac{p^j}{1 - p^j}\right)^n \quad (2.13)$$

and

$$E(Y^n) = \sum_{i=0}^{\infty} \omega_i(a, b) \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\alpha(a+i)}{j} \sum_{t=1}^n S(n, t) t! \left(\frac{p^j}{1 - p^j}\right)^t, \quad (2.14)$$

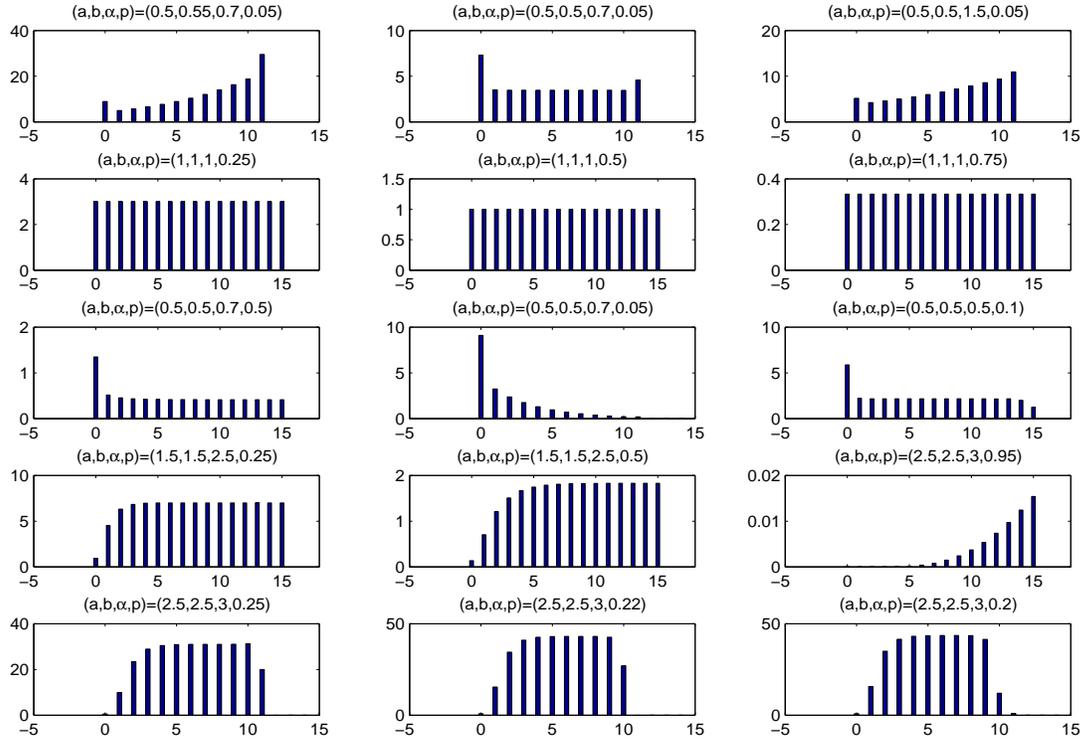


Figure 1: Hazard rate function plots of the DBGE distribution for selected parameters values.

where

$$S(n, t) = \frac{1}{t!} \sum_{i=0}^{t-1} (-1)^i \binom{t}{i} (t-i)^n,$$

is the Stirling number of the second kind. In particular, the mean and variance of the distribution are given, respectively, by

$$E(Y) = \sum_{i=0}^{\infty} \omega_i(a, b) \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\alpha(a+i)}{j} \frac{p^j}{1-p^j}, \tag{2.15}$$

and

$$Var(Y) = \sum_{i=0}^{\infty} \omega_i(a, b) \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\alpha(a+i)}{j} \frac{p^j(1+p^j)}{(1-p^j)^2} - \{E(Y)\}^2. \tag{2.16}$$

Here, we consider some special cases of *DBGE* distributions which are interesting and useful for their own sake.

1) If $a = b = 1$, then $DBGE(a, b, \alpha, p)$ distribution reduces to $DGE_2(\alpha, p)$ distribution. In addition, if $\alpha = 1$, then $DBGE(1, 1, \alpha, p)$ distribution is just the geometric distribution; see Nekoukhou et al. (2013a).

2) If $a = 1$, then $DBGE(a, b, \alpha, p)$ distribution reduces to a discrete version of the double generalized exponential distribution, $DDGE(b, \alpha, p)$, of Barreto-Souza et al. (2010). This has not been introduced in the literature yet.

3) If $\alpha = 1$, as we know, BGE distribution reduces to BE distribution of Nadarajah and Kotz (2006). Hence, for $\alpha = 1$, DBGE distribution gives a discrete counterpart of BE distribution. It should be noted that, recently, Nekoukhou et al. (2013b) considered another discrete version of the BE distribution of Nadarajah and Kotz (2006) using the relation

$$P_y = \frac{f(y)}{\sum_{t=1}^{\infty} f(t)}, \quad y = 1, 2, \dots,$$

where f is the pdf of BE distribution. These two discrete analogues of BE distribution have different structures.

3 Application

In this section, the DBGE model will be examined for a real data set which is given by Karlis and Xekalaki (2001) on the numbers of fires in Greece for the period from 1 July 1998 to 31 August of the same year. This data set consists of 123 observation and has been presented in Table 1. Only fires in forest districts are considered. Bakouch et al. (2014) considered these data to indicate the potentiality of discrete Lindley (DL) distribution in data modeling and compared it with Poisson, geometric, discrete Weibull (DW) and discrete gamma (DG) distributions.

Table 1 Numbers of fires in Greece.

Numbers	0	1	2	3	4	5	6	7	8	9	10	11	12	15	16	20	43
Frequency	16	13	14	9	11	13	8	4	9	6	3	4	6	4	1	1	1

Here, we compare the DBGE, DBE, DDGE and DGE_2 models with these discrete distributions. In addition, because of the over dispersion phenomena in the data set, $\bar{x} = 5.3984$ and $s^2 = 30.0449$, the

negative binomial (NB) distribution is also compared with the others. Maximum likelihood method is used to obtain the estimates of the parameters of the proposed new distributions (DBGE, DBE and DDGE). Comparing the DBGE model with the other models is performed by using the Akaike information criterion (AIC) and Kolmogrov-Smirnov (K-S) test statistic. Table 2 indicates the fitting computations which consists of the MLEs, AICs and K-S test statistics determined by the fitting models.

According to the AICs and K-S test statistics values in Table 2, it seems that DBGE model gives a better fit than DL, DW, DG, NB, Poisson, geometric and also DBE, DDGE and DGE_2 distributions. In addition, the empirical and probability-probability (p-p) plots of DBGE distribution, given in Figure 2, confirm a satisfactory fit for this new discrete model.

One can construct approximate confidence intervals for the parameters of DBGE model. Indeed, such confidence intervals are attained by means of the asymptotic covariance matrix of the MLEs of DBGE parameters when the Newton-Raphson procedure converges. For instance, 95% asymptotic confidence intervals for DBGE parameters are calculated as $a \in (1.2668 \mp 0.8254)$, $b \in (3.0100 \mp 0.7289)$, $\alpha \in (1.1030 \mp 0.5120)$ and $p \in (0.9354 \mp 0.1592)$.

Table 2 Summary.

Models	MLEs	AIC	K-S statistic
DBGE	$(\hat{a}, \hat{b}, \hat{\alpha}, \hat{p}) = (1.2668, 3.0100, 1.1030, 0.9593)$	121.7440	0.0815
DBE	$(\hat{a}, \hat{b}, \hat{p}) = (2.3176, 1.2883, 0.9031)$	125.8864	0.1666
DDGE	$(\hat{b}, \hat{\alpha}, \hat{p}) = (3.3213, 0.7431, 0.9463)$	142.9386	0.5226
DGE_2	$(\hat{\alpha}, \hat{p}) = (1.2548, 0.8225)$	130.3965	0.3386
NB	$(\hat{r}, \hat{p}) = (1.3360, 0.1984)$	126.9362	0.3350
DL	$(\hat{\theta}, \hat{p}) = (0.3090, 0.7343)$	131.5166	0.1122
DW	$(\hat{p}, \hat{\alpha}) = (0.8789, 1.1308)$	127.5752	0.3351
DG	$(\hat{\alpha}, \hat{\beta}) = (0.7525, 0.1543)$	137.1818	0.2683
Poisson	$\hat{\lambda} = 5.3984$	253.5164	0.2547
Geometric	$\hat{p} = 0.8437$	129.5689	0.1636

Conclusion

We introduced a new discrete distribution, called discrete beta generalized exponential (DBGE) distribution, motivated by the fact that the distribution provides greater flexibility to analyze more complex

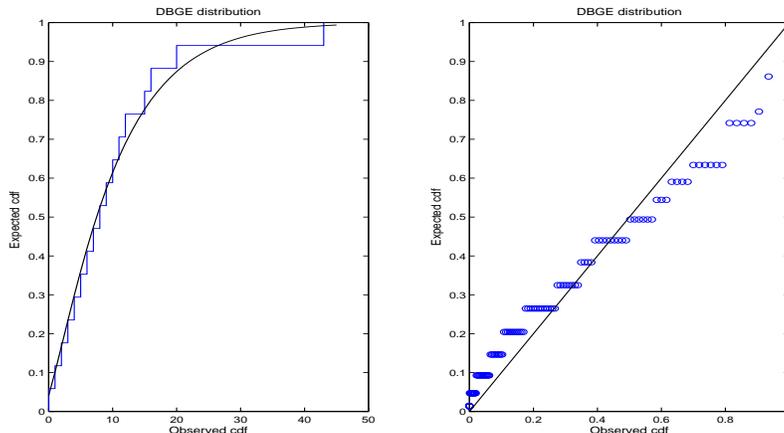


Figure 2: The empirical cdf (left) and probability-probability (right) plots of DBGE distribution.

situations. Indeed, the hazard rate function of the new introduced model can be bathtub, unimodal, increasing, decreasing and constant. That is, the DBGE distribution can be used as an improved model for failure data in the discrete case.

References

- Bakouch, H.S., Aghababaei Jazi, M. and Nadarajah, S. (2014). A new discrete distribution. *Statistics*, 48(1):200-240.
- Barreto-Souza, W., Santos, A.H.S. and Cordeiro, G.M. (2010). The beta generalized exponential distribution. *J. Stat. Comput. Simul.*, 80:159-172.
- Bidram, H., Behboodian, J. and Towhidi, M. (2013). The beta Weibull-geometric Distribution. *J. Stat. Comput. Simul.*, 83(1):52-67.
- Cordeiro, G.M., Gomes, A.E., Silva, C.Q. and Ortega, E.M.M. (2013). The beta exponentiated Weibull distribution. *J. Stat. Comput. Simul.*, 83(1):114-138.
- Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Comm. Stat. Theo. Meth.*, 31:497-512.

- Gupta, R.D. and Kundu, D. (1999). Generalized exponential distributions. *Aust. NZ. J. Stat.*, 41(2):173-188.
- Karlis, D. and Xekalaki, E. (2001). On some discrete valued time series models based on mixtures and thinning, in Proceedings of the Fifth Hellenic-European Conference on Computer Mathematics and Its Applications, E.A. Lipitakis, ed., 872-877.
- Mahmoudi, E. (2011). The beta generalized Pareto distribution with application to lifetime data. *Math. Comput. Simul.*, 81:2414-2430.
- Nadarajah, S. and Kotz, S. (2006). The beta exponential distribution. *Reli. Eng. Syst. Saf.*, 91:689-697.
- Nekoukhou, V., Alamatsaz, M.H. and Bidram, H. (2013a). Discrete generalized exponential distribution of a second type. *Statistics*, 47(4):876-887.
- Nekoukhou, V., Alamatsaz, M.H., Bidram, H. and Aghajani, A.H. (2013b). Discrete Beta-Exponential Distribution. *Comm. Stat. Theo. Meth.*, DOI: 10.1080/03610926.2013.773348.



Some Results on APND Random Variables

H.R.Nili Sani¹, M. Amini², A.Bozorgnia³

¹ Dept. of Stat., University of Birjand, Birjand-Iran.

²Dept. of Stat., University of Ferdowsi, Mashhad-Iran.

³Dept. of Stat., University of Khayyan, Mashhad-Iran.

Abstract: Asymptotically pairwise negatively dependent random variables were introduced by Nili Sani et.al(2015). This class includes independent random variables(hencefort r.v.'s) and pairwise negatively dependence r.v.'s. In this note, we discuss the asymptotic behavior of the tail of $X_1 + X_2$ that X_1, X_2 are APND r.v.'s and extend some results of Rangbar et.al.(2008) and (2009).The limit theorems for the new class of r.v.'s are also proved and some results of Beak and Park (2010) and Soo Hak Sung(2014) are extended to this class conveniently.

Keywords Max-sum equivalence, Negatively dependent r.v.'s, Asymptotically pairwise negatively dependent r.v.'s, Complete convergence, Fundamental maximal inequality.

Mathematics Subject Classification (2010): 60F15.

1 Introduction

Lehmann (1966) introduced various concepts of positive (negative) dependence for two random variables. Since then, dependent random variables have captured the attention of many researchers and they have been extensively studied in the literature. Although various results in probability and statistics have been derived under the assumption that some underlying random variables have the negative dependence property (for details and references, see Lehmann(1966), Joag-Dev and Proschan (1983), Joe (1996), Bozorgnia et. al.(1996) and Amini(2000)), the positive dependent concept is more practicable. In the next section, we introduce a new class of dependent r.v.'s and provide some basic properties of it. This class is bigger than of negative dependent r.v.'s class. In the last section, we discuss the limit behavior of the weighted sums of arrays of new r.v.'s.

³Speaker's Name :Nili Sani nilisanil@birjand.ac.irr

2 Preliminaries

Definition 2.1. *The random variables X_i and X_j are said to be asymptotic negatively dependent if the following inequalities hold,*

$$P(X_i > x_i, X_j > x_j) \leq (1 + q(|i - j|))P(X_i > x_i)P(X_j > x_j) \quad (2.1)$$

and

$$P(X_i \leq x_i, X_j \leq x_j) \leq (1 + q(|i - j|))P(X_i \leq x_i)P(X_j \leq x_j) \quad (2.2)$$

for all $x_i, x_j \in R, i \neq j$, where $q(n) = o(n^{-w}), w > 0$.

Similarly, (X_i, X_j) is asymptotic positively dependent if (2.1) and (2.2) hold with the inequalities sign reversed.

The random variables X_1, X_2, \dots are said to be *asymptotic pairwise negatively dependent* (APND) if (X_i, X_j) is asymptotic negatively dependent for every $i \neq j, i, j \geq 1$.

If for $i, j, q(|i - j|) = 0$, then the random variables X_i and X_j are negatively dependent (Lehmann (1966)). A special subclass of APND r.v.'s is the pairwise negatively dependent r.v.'s studied by NiliSani et.al.(2005).

Lemma 2.2. (NiliSani et.al.(2015)) *Let $\{X_n\}$ be a sequence of APND r.v.'s and $\{f_n\}$ be a sequence of concordinate functions.*

(i). *The set of inequalities (2.1) and (2.2) are equivalent to that obtained by the replacing one or more than one of inequalities $X_i \leq x_i, X_j \leq x_j, X_i > x_i$ or $X_j > x_j$ by the corresponding $X_i < x_i, X_j < x_j, X_i \geq x_i$ or $X_j \geq x_j$.*

(ii) *$\{f_n(x_n)\}$ is a sequence of APND r.v.'s with the initial coefficients $q(n)$.*

(iii) *$P(X_i \geq x_i, X_j < x_j) \geq P(X_i \geq x_i)[1 - (1 + q|i - j|)P(X_j \geq x_j)]$*

(iv) *$P(X_i \leq x_i, X_j > x_j) \geq P(X_i \leq x_i)[1 - (1 + q|i - j|)P(X_j \leq x_j)]$*

Lemma 2.3. (NiliSani et.al.(2015)) *Let X_i, X_j be two APND r.v.'s. Then*

$$Cov(X_i, X_j) \leq 4q(|i - j|)E|X_i|E|X_j|,$$

provided the expectations on the left hand side exist.

Lemma 2.4. (NiliSani et.al.(2015)) Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be independent pairs of APND r.v.'s with joint distributions F_1, F_2, \dots, F_n and coefficient q_1, q_2, \dots, q_n . Let f and g be concordant Borel functions, $X = f(X_1, \dots, X_n)$ and $Y = g(Y_1, \dots, Y_n)$, then (X, Y) are APND.

Example 2.5. Farlie-Gumbel-Morgenstern (FGM) is a famous family of bivariate distributions (for details, references and some applications, see Long and Krzysztofowicz(1992) and Drouet Mari and Kotz(2004)). The joint distribution function is as follows

$$F(x, y) = F_1(x)F_2(y)(1 + \alpha(1 - F_1(x))(1 - F_2(y))), \quad |\alpha| < 1. \tag{2.3}$$

(2.3) is equivalent to

$$P(X > x, Y > y) = P(X > x)P(Y > y)(1 + \alpha F_1(x)F_2(y)).$$

FGM family is a subset of APND random variables.

The result below extends the result of Ranjbar et.al. (2008) from weakly negative dependent to APND r.v.'s.

Lemma 2.6. If two APND random variables X_1 and X_2 have distributions F_1 and F_2 respectively, then $F_1 \sim_M F_2$, is equivalent to

$$P(X_1 + X_2 \geq x) \sim P(\max\{X_1, X_2\} \geq x).$$

Proof. $P(X_1 + X_2 > x) \geq P(\max(X_1, X_2) > x) \geq P(X_1 > x) + P(X_2 > x) - P(X_1 > x)P(X_2 > x)$.
If $F_1 \sim_M F_2$, then

$$1 \geq \lim \frac{P(\max(X_1, X_2) > x)}{P(X_1 + X_2 > x)} \geq 1 - \lim \frac{c\overline{F_1}(x)\overline{F_2}(x)}{P(X_1 + X_2 > x)} = 1.$$

The proof $P(X_1 + X_2 > x) \sim P(\max(X_1, X_2) > x) \Rightarrow F_1 \sim_M F_2$ is obvious. □

Lemma 2.7. Let X_i and X_j two nonnegative APND random variables with distributions F and G respectively, then

$$\frac{P(X_i + X_j > x)}{P(X_i > x) + P(X_j > x)} \geq 1 + o(1)$$

Proof. For any positive x we have

$$\begin{aligned} P(X_i + X_j > x) &= P(X_i > x) + P(X_j > x) - P((X_i > x), (X_j > x)) \\ &\quad + P(X_i + X_j > x, X_i < x, X_j < x) \\ &\geq P(X_i > x) + P(X_j > x) - P(X_i > x)P(X_j > x)(1 + q(|i - j|)) \end{aligned}$$

then

$$\frac{P(X_i + X_j > x)}{P(X_i > x) + P(X_j > x)} \geq 1 + o(1).$$

□

Lemma 2.8. Let X_i and X_j two APND random variables with distributions F and G respectively, then

$$\frac{P(X_i + X_j > x)}{P(X_i > x) + P(X_j > x)} \geq 1 + o(1)$$

Proof. For any positive x we have

$$\begin{aligned} P(X_i + X_j > x) &= P(X_i^+ + X_j^+ - (X_i^- + X_j^-) > x) \\ &= P(X_i + X_j > x, X_i > 0, X_j > 0) + P(X_i + X_j > x, X_i < 0, X_j < 0) \\ &\quad + P(X_i + X_j > x, X_i > 0, X_j < 0) + P(X_i + X_j > x, X_i < 0, X_j > 0) \\ &\geq P(X_i^+ + X_j^+ > x) \\ &\geq P(X_i^+ > x) + P(X_j^+ > x) - P(X_i^+ > x)P(X_j^+ > x)(1 + q(|i - j|)) \end{aligned}$$

then

$$\frac{P(X_i + X_j > x)}{P(X_i > x) + P(X_j > x)} \geq 1 - q(|i - j|) \frac{P(X_i^+ > x) + P(X_j^+ > x)}{P(X_i > x) + P(X_j > x)} = 1 + o(1)$$

□

Theorem 2.9. Let X_1, \dots, X_n be a sequence r.v.'s that satisfy in (1) for every $1 \leq i, j \leq n$, then

$$P(\max_{1 \leq i \leq n} X_i > x) \sim \sum_{i=1}^n P(X_i > x) \quad \text{as } x \rightarrow \infty$$

The proof, similar to that of Theorem 4 in Ranjbar et.al. (2008), is omitted.

3 SLLN for APND r.v.'s

Theorem 3.1. (Fundamental maximal inequality). Let $\{X_n\}$ be a sequence of APND r.v.'s with coefficients $q(m)$, $w > 2$, such that with $E(X_n) \uparrow$. Then

$$E[\max_{1 \leq i \leq m} S_i^2] \leq \left[\frac{\log(4n)}{\log 2} \right]^2 \sum_{i=1}^n E(X_i^2) \quad \text{for each } n \geq 1$$

Proof. The proof is similar to that of theorem 2.3.1 in Stout (1974). The first, we suppose that $n = 2^k$, $k = 6$. Let $X_{r,s} = \sum_{i=r+1}^s X_i$, $0 \leq r < s \leq 2^6$. Consider the following collections of $X_{r,s}$:

$$\{X_{0,64}\}, \{X_{0,32}, X_{32,64}\}, \dots, \{X_{0,1}, \dots, X_{63,64}\}$$

There are $k + 1 = 7$ collections. Represent the expansion S_i by $S_i = \sum_{j=1}^h X_{i_{j-1}, i_j}$, $h \leq 7$. Then $S_i^2 \leq 7 \sum_{j=1}^h (X_{i_{j-1}, i_j})^2$ and

$$E(\max_{1 \leq i \leq 2^6} S_i^2) \leq 7.7 \sum_{j=1}^{2^6} E(X_i^2) \left(1 + \sum_{r=1}^{64-i} q(r) \right)$$

follows by the Cauchy-Schwarz inequality and Lemma 2.

Using a suitable notation, the above argument generalized to arbitrary $k \geq 1$ yields

$$E(\max_{1 \leq i \leq 2^k} S_i^2) \leq (k + 1)^2 \sum_{j=1}^{2^k} E(X_j^2) (1 + \sum_{r=1}^{n-i} q(r))$$

and in general

$$E(\max_{1 \leq i \leq n} S_i^2) \leq (k + 1)^2 \sum_{i=1}^n E(X_i^2) (1 + \sum_{r=1}^{n-i} q(r)).$$

Since $2^{k-1} < n$ implies $(k + 1)^2 \leq [\log(4n)/\log 2]^2$, the desired result follows. □

Remark 3.2. : If $n > a^2$, then $\log 2$ can be replaced by $\log a$.

Theorem 3.3. Let $\{X_{ni}\}$ be an array of rowwise APND r.v.'s with $E(X_{ni}) = 0$ and common coefficients $q(m), w > 2$, and for some random variable X and constant $c > 0$,

$$P(|X_{ni}| > x) \leq cP(|X| > x) \tag{3.1}$$

for all n, i and $x \geq 0$. Suppose that W is non-negative r.v. with distribution function G such that

$$E(W^{-2}) < \infty. \tag{3.2}$$

Moreover, suppose that $\beta \geq -1$ and $\{a_{ni}\}$ is an array of constants such that

$$\sup_{i \geq 1} |a_{ni}| = O(n^{-3r/4}) \quad \text{for some } r > 0 \tag{3.3}$$

and

$$\sum_{i=1}^{\infty} |a_{ni}| = O(n^\alpha) \quad \text{for some } \alpha \in [0, r). \tag{3.4}$$

If $\alpha + \beta > -1$ and there exists some $1 < \delta \leq \frac{4}{3}(1 + \beta)$ such that

$$\beta - \frac{3}{4}r(\delta - (\frac{4\alpha}{3r} + 1)) < -1, \tag{3.5}$$

and

$$E(|X|^\theta) < \infty \tag{3.6}$$

where $2 > \theta > \frac{4}{3r}(1 + \beta + \alpha) + 1$, then

$$\sum_{n=1}^{\infty} n^\beta \int_0^\infty P(|\sum_{i=1}^{\infty} a_{ni} X_{ni}| > w) dG(w) < \infty. \tag{3.7}$$

Proof. The proof is the same as that of Theorem 2 of NiliSani (2015). Note that $a_{ni} = a_{ni}^+ - a_{ni}^-$. Thus, to prove (3.7), it suffices to show that, for every $w > 0$
 $\sum_{n=1}^{\infty} n^{\beta} \int_0^{\infty} P(|\sum a_{ni}^+ X_{ni}| > \frac{w}{2})dG(w) < \infty$ and $\sum_{n=1}^{\infty} n^{\beta} \int_0^{\infty} P(|\sum a_{ni}^- X_{ni}| > \frac{w}{2})dG(w) < \infty$.
 So, we may assume, without loss of generality, that $a_{ni} > 0$. Let

$$X'_{ni} = -I_{(a_{ni}X_{ni} \leq -1)} + a_{ni}X_{ni}I_{(|a_{ni}X_{ni}| < 1)} + I_{(a_{ni}X_{ni} \geq 1)}.$$

Then

$$\sum_{n=1}^{\infty} n^{\beta} \int_0^{\infty} P(|\sum_{i=1}^{\infty} a_{ni}X_{ni}| > \frac{w}{2})dG(w) \leq \sum_{n=1}^{\infty} n^{\beta} \int_0^{\infty} P(|\sum_{(|a_{ni}X_{ni}| \geq 1)} a_{ni}X_{ni}| > \frac{w}{4})dG(w) + \sum_{n=1}^{\infty} n^{\beta} \int_0^{\infty} P(|\sum_{i=1}^{\infty} X'_{ni}| > w/4)dG(w) = I_1 + I_2.$$

Hence, it suffices to show that $I_1 < \infty$ and $I_2 < \infty$. We assume, without loss of generality, that

$$\sup_{i \geq 1} a_{ni} = n^{-\frac{3r}{4}}, \quad \sum_{i=1}^{\infty} a_{ni} = n^{\alpha}, \quad a_{ni} \leq 1$$

Let

$$I_{nk} = \{i : (nk)^{\frac{3r}{4}} \leq |b_{ni}| = |\frac{1}{a_{ni}}| < (n(k+1))^{\frac{3r}{4}}\}, \quad \sum_{j=1}^k \#I_{nj} =: \#Ank$$

for all $k \geq 1$ and $n \geq 1$, then $\bigcup_{j=1}^k I_{nj} = N$ for all $n \geq 1$. It is easy to see that

$$\sum_{j=1}^k \#I_{nj} =: \#Ank \leq n^{\alpha}(n(k+1))^{\frac{3r}{4}}. \tag{3.8}$$

We are going to show that $I_1 < \infty$. Using (3.4) we obtain

$$I_1 = \sum_{n=1}^{\infty} n^{\beta} \int_0^{\infty} P(|\sum_{(|a_{ni}X_{ni}| \geq 1)} a_{ni}X_{ni}| > \frac{w}{4})dG(w) \leq \sum_{n=1}^{\infty} n^{\beta} \int_0^{\infty} P(|a_{ni}X_{ni}| \geq 1)dG(w) \leq c \sum_{n=1}^{\infty} n^{\beta} P(|X| \geq \frac{1}{|a_{ni}|}) \leq c \sum_{n=1}^{\infty} n^{\beta} \frac{E(|X|^{\theta})}{|a_{ni}|^{\theta}} < \infty.$$

Thus to prove $I_2 < \infty$, we need only to prove that

$$\sum_{n=1}^{\infty} n^{\beta} (E(\sum X'_{ni}))^2 < \infty \tag{3.9}$$

and

$$\sum_{n=1}^{\infty} n^{\beta} \sigma^2(\sum X'_{ni}) < \infty. \tag{3.10}$$

Decomposing the left-hand side (3.9) into the two parts and by $E(X_{ni}) = 0$, we see that

$$|E(\sum_{i=1}^{\infty} X'_{ni})| = |E(\sum_{i=1}^{\infty} (X'_{ni}I_{|X'_{ni}| < 1})) + E(\sum_{i=1}^{\infty} (X'_{ni}I_{|X'_{ni}| \geq 1}))|$$

$$\leq c \sum_{i=1}^{\infty} E(a_{ni}|X|I_{|a_{ni}X|\geq 1}) \leq c \sum_{i=1}^{\infty} a_{ni} E(\frac{|X|^{\theta}}{|X|^{\theta-1}}I_{|X|\geq a_{ni}}) \leq cn^{-1-\beta},$$

therefore

$$\sum_{n=1}^{\infty} n^{\beta} (E(\sum X'_{ni}))^2 < \infty.$$

Now, we are going to show that the validity of (3.10),

$$\sum_{n=1}^{\infty} n^{\beta} Var(\sum_{i=1}^{\infty} X'_{ni}) \leq c[\sum_{n=1}^{\infty} n^{\beta} [\sum_{i=1}^{\infty} EX'_{ni}{}^2 + \sum_{i=1, j>i}^{\infty} (q(|i-j|))E(|X'_{ni}|)E(|X'_{nj}|)].$$

In order to prove that each terms of RHS is finite, we need the following inequalities

$$\begin{aligned} E(|X'_{ni}|)E(|X'_{nj}|) &\leq [E(I_{(|a_{ni}X_{ni}|\geq 1)}) + E(|a_{ni}X_{ni}|I_{(0<a_{ni}X_{ni}<1)})][E(I_{(|a_{nj}X_{nj}|\geq 1)}) \\ &\quad + E(|a_{nj}X_{nj}|I_{(0<a_{nj}X_{nj}<1)})] \\ &\leq 4E(|a_{ni}X_{ni}|I_{(|a_{ni}X_{ni}|\geq 1)})E(|a_{nj}X_{nj}|I_{(|a_{nj}X_{nj}|\geq 1)}) \\ &\leq ca_{ni}a_{nj}E(|X|I_{(|X|\geq \frac{1}{a_{ni}})})E(|X|I_{(|X|\geq \frac{1}{a_{nj}})}) \leq ca_{ni}a_{nj}n^{-2(\theta-1)\frac{3r}{4}}. \end{aligned}$$

Since

$$\begin{aligned} \sum_{n=1}^{\infty} n^{\beta} \sum_{i=1, j<i}^{\infty} (q(i-j))E(|X'_{ni}|)E(|X'_{nj}|) &\leq c \sum_{n=1}^{\infty} n^{\beta} \sum_{i=1, j>i}^{\infty} q(|i-j|)n^{-2(\theta)\frac{3r}{4}} \\ &\leq c \sum_{n=1}^{\infty} n^{\beta-2\theta\frac{3}{4r}} \sum_{i=1, j>i}^{\infty} q(i) < \infty, \end{aligned}$$

it is sufficient to show that the first term of RHS,

$$\sum_{n=1}^{\infty} n^{\beta} \sum_{i=1}^{\infty} EX'_{ni}{}^2 = \sum_{n=1}^{\infty} n^{\beta} \sum_{i=1}^{\infty} P(|a_{ni}X_{ni}| \geq 1) + \sum_{n=1}^{\infty} n^{\beta} \sum_{i=1}^{\infty} E((a_{ni}X_{ni})^2 I_{(|a_{ni}X_{ni}|<1)})$$

is finite. Note that

$$\sum_{n=1}^{\infty} n^{\beta} \sum_{i=1}^{\infty} P(|a_{ni}X_{ni}| \geq 1) \leq \sum_{n=1}^{\infty} n^{\beta} \sum_{i=1}^{\infty} P(|X| \geq \frac{1}{a_{ni}}) \leq c \sum_{n=1}^{\infty} n^{\beta} \sum_{i=1}^{\infty} a_{ni}^{\theta} < \infty.$$

By (3.1) we have

$$\begin{aligned} E((a_{ni}X_{ni})^2 I_{(|a_{ni}X_{ni}|<1)}) &= \int_0^{\infty} P((a_{ni}X_{ni})^2 I_{(|a_{ni}X_{ni}|<1)} \geq x) dx \\ &= \int_0^{\frac{1}{a_{ni}}} 2a_{ni}^2 P(y \leq X_{ni} < \frac{1}{a_{ni}}) y dy \leq ca_{ni}^2 \int_0^{\frac{1}{a_{ni}}} y \frac{E(|X|^{\theta})}{y^{\theta}} dy \leq ca_{ni}^{\theta} \end{aligned}$$

and, hence

$$\sum_{n=1}^{\infty} n^{\beta} \sum_{i=1}^{\infty} EX'_{ni}{}^2 \leq c \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} n^{\beta} a_{ni}^{\theta} = c \sum_{n=1}^{\infty} n^{\beta} a_{ni}^{\theta} \sum_{i=1}^{\infty} a_{ni} < \infty,$$

the last inequality follows from (3.3), (3.4) and (3.6). □

References

Amini, M. (2000). *Ph.D Thesis.*, Ferdowsi University., Mashhad-Iran..

Beak, J.-II and Park, S.T. (2010). Convergence of weighted sums for arrays of negatively dependent random variables and its applications, *J. of Statistical Plan. and infer.*, **140**, 2461-2469.

Borzorgnia, A. , Patterson, R.F. and Taylor, R. L. (1996). Limit theorems for dependent random variables. *World Congress Nonlinear Analysts* **92** 1639-1650.

- Drouet Mari,K. and Kotz,S.(2001). *Correlation and Dependence*, Imperial College Press.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.
- Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications, *Ann. Statist.*, **11**, 286-295.
- Hsu, P.L. and Robbins, H. (1947). Complete convergence and the law of large numbers, *Proc. Nat. Acad. Soc., U.S.A.*, **33**, 25-31.
- Lehmann, E.L. (1966). Some concept of dependence, *Ann. Math. Statist.*, **37**, 1137-1153.
- Long, D. and Krzysztofowicz, R. (1992). Farlie-Gumbel-Morgenstern bivariate densities: Are they applicable in hydrology?, *Stochastic Hydrology and Hydraulics*, **6**, 47-54.
- Nili Sani, H.R., Amini, M. and Bozorgnia, A.(2005). Strong Laws for Weighted Sums of Negative Dependent Random Variables, *J. of Sci., I.R.I.*, **16(3)**, 261-265.
- Nili Sani, H.R., Amini, M. and Bozorgnia, A.(2015). Complete convergence for weighted sums of arrays of APND r.v.'s, *Submitted*.
- Ranjbar, V., Amini,M. and Bozorgnia, A. (2008). Asymptotic behaviour of weighted sums of weakly negative dependent random variables, *J. of Sci. I.R.I.*, **19(4)**, 357-363.
- Sung, S.H. (2001). Strong laws for weighted sums of i.i.d r.v.'s *Statist. Probab. Lett.* **52**, 413-419.
- Taylor, R.L. and Hu,T.C. (1987). Strong laws of large numbers for arrays of rowwise independent random elements. *Internat. J. Math.& Math.Sci.*, **10(4)**, 805-814.



Some Asymptotic Results on Improved LASSO Estimators

M. Norouzirad¹, M. Arashi², A.K.Md. Ehsanes Saleh³

^{1,2}Department of Statistics, School of Mathematical Sciences, University of Shahrood,
Shahrood, Iran,

³School of Mathematics and Statistics, Carleton University, Ottawa, Canada

Abstract: In the context of regression modeling, when the number of predictors are relatively large with respect to sample size, the least absolute shrinkage and selection operator (LASSO), adds some level of sparsity to the estimation of parameter vector to improve the prediction. When the coefficients are subjected to lie in a subspace hypothesis, we propose to use a restricted LASSO estimator along with its preliminary test and shrinkage versions. Asymptotic distributions with properties of the newly defined LASSO-type estimators are derived and some graphs are depicted to show the performance of asymptotic relative efficiencies.

Keywords LASSO, Asymptotic distributional risk, Asymptotic relative efficiency, Preliminary test LASSO, Shrinkage LASSO.

Mathematics Subject Classification (2010): 62F12, 62J07.

1 Introduction

Consider the linear regression model

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.1)$$

where \mathbf{y} is an n -vector of observations on the response variable, X is a known $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a p -vector of unknown regression coefficients, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)^T$ such that ε_i s are i.i.d random variables with mean 0 and variance σ^2 . Without loss of generality, we will assume the covariates are centered to have mean 0.

¹Speaker's Name: norouzirad@shahroodut.ac.ir

We estimate β by minimizing the L_1 -penalized least squares criterion

$$\sum_{i=1}^n (y_i - \mathbf{x}_i^T \phi)^2 + \lambda_n \sum_{j=1}^p |\phi_j|, \quad (1.2)$$

for a given λ_n . Such estimator were called ‘‘LASSO’’ by Tibshirani (1996).

A ‘‘LASSO’’ is usually recognized as a loop of rope that is designed to be thrown around a target and tighten when pulled; a well-known tool of the American cowboy. In statistical context, LASSO (Least Absolute shrinkage and Selection Operator) is fittingly being used as a metaphor of L_1 constraint applied to linear model.

For a given λ_n , we denote the LASSO estimator by $\tilde{\beta}_n^L$. If $\lambda_n = 0$, it corresponds to the ordinary least squares (OLS) estimator that will be denoted by $\tilde{\beta}_n^{(0)} = (X^T X)^{-1} X^T \mathbf{y}$.

We assume that the following regularity conditions is hold for the design,

$$C_n = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \rightarrow C \quad (1.3)$$

where C is a nonnegative definite matrix and

$$\frac{1}{n} \max_{1 \leq i \leq n} \mathbf{x}_i^T \mathbf{x}_i \rightarrow 0 \quad (1.4)$$

Usually, the covariates are scaled; Thus, the diagonal elements of C_n (and hence those of C) are all identically 1.

Under conditions (1.3) and (1.4) (with C nonsingular), it is well-known that the OLS estimator is consistent and that

$$\sqrt{n} \left(\tilde{\beta}_n^{(0)} - \beta \right) \rightarrow^D \mathcal{N}_p(\mathbf{0}, \sigma^2 C^{-1}).$$

In fact, conditions (??) and (??) can be weakened considerably without losing asymptotic normality of the OLS estimator Srivastava (1971).

By the analogy of Knight and Fu (2000) in order to show consistency of $\tilde{\beta}_n^L$, we will define the (random) function

$$Z_n(\phi) = \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \phi)^2 + \frac{\lambda_n}{n} \sum_{j=1}^p |\phi_j|, \quad (1.5)$$

which is minimized at $\phi = \tilde{\beta}_n^L$.

Theorem 1.1 (Knight and Fu , 2000). (i) Suppose that C in (1.3) is nonsingular,

(i) If $\lambda_n/n \rightarrow \lambda_0 \geq 0$, then $\tilde{\beta}_n^L \xrightarrow{P} \operatorname{argmin}(Z)$ where $Z(\phi) = (\phi - \beta)^T C (\phi - \beta) + \lambda_0 \sum_{j=1}^p |\phi_j|$.

(ii) If $\lambda_n/\sqrt{n} \rightarrow \lambda_0 \geq 0$, then

$$\sqrt{n}(\tilde{\beta}_n^L - \beta) \xrightarrow{D} \operatorname{argmin}(V)$$

where $V(\mathbf{u}) = -2\mathbf{u}^T W + \mathbf{u}^T C \mathbf{u} + \lambda_0 \sum_{j=1}^p [u_j \operatorname{sgn}(\beta_j) I(\beta_j \neq 0) + |u_j| I(\beta_j = 0) - |\beta_j|]$, and $W \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 C)$.

Corollary 1.2. (i) If $\lambda_n = o(n)$, then $\tilde{\beta}_n^L$ is consistent.

(ii) If $\lambda_n = o(\sqrt{n})$, then $\sqrt{n}(\tilde{\beta}_n^L - \beta) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, \sigma^2 C)$.

Even though $\lambda_n = o(n)$ is sufficient for consistency, we require that λ_n grow more slowly for \sqrt{n} -consistency of LASSO estimator. However, if λ_n grows too slowly then $\sqrt{n}(\tilde{\beta}_n^L - \beta)$ will have the same limiting distribution as $\sqrt{n}(\tilde{\beta}_n^{(0)} - \beta)$. In fact, the rate of growth of λ_n needed to get an “interesting” limiting distribution. The next theorem indicated that we need $\lambda_n = O(\sqrt{n})$ for \sqrt{n} -consistency (Knight and Fu , 2000).

1.1 Improved LASSO Estimators

Up to the point, it was assumed that the level of information had depend on the sample, assuming no non-sample effect in estimation procedure. In this sense, we denote a LASSO estimator of β by $\tilde{\beta}_n^L$ and term it the unrestricted LASSO estimator (ULE).

However, in some situations it is possible to have some non-sample information (a priori restriction on the parameters), usually subjected to the model as constraints.

A set of q linear restrictions on the vector β can be written as $H\beta = \mathbf{h}$. Or we can suppose that our model is subjected to lie in the linear sub-space restriction

$$H\beta = \mathbf{h}, \tag{1.6}$$

where H is a $q \times p$ ($q \leq p$) matrix of known elements, with q being the number of linear restriction to test, and \mathbf{h} is a $q \times 1$ vector of known elements. The rank of H is q , which implies that the restrictions are linearly independent. This restriction may be (a) a fact known from theoretical or experimental considerations, (b) a hypothesis that may have to be tested or (c) an artificially imposed condition to reduce or eliminate redundancy in the description of the model (see Arashi and Nadarajah , 2014).

Under the restrain model (1.6), the estimator is called as the restricted LASSO estimator (RLE), denoted by $\hat{\beta}_n^{RL}$. By the analogy of OLS estimator of β , subject to the restriction $H\beta = \mathbf{h}$, we propose the

$$\hat{\beta}_n^{RL} = \tilde{\beta}_n^L - C_n^{-1}H^T(HC_n^{-1}H^T)^{-1}(H\tilde{\beta}_n^L - \mathbf{h}). \quad (1.7)$$

When (1.6) is satisfied, $\hat{\beta}_n^{RL}$ has smaller asymptotic risk than $\tilde{\beta}_n^L$; However, for $H\beta \neq \mathbf{h}$, $\hat{\beta}_n^{RL}$ may be biased and inconsistent in many cases. For this reason, it is plausible to consider a preliminary test LASSO estimator (PTLE) by taking $\tilde{\beta}_n^L$ or $\hat{\beta}_n^{RL}$ according to acceptance or rejection of the null hypothesis H_0 . We denote it by $\hat{\beta}_n^{PTL}$. In order to propose a PTLE, we consider the preliminary test for the null hypothesis $\mathcal{H}_o : H\beta = \mathbf{h}$ based on

$$T_n^L = \frac{(H\tilde{\beta}_n^{(0)} - \beta)^T(HC_n^{-1}H^T)^{-1}(H\tilde{\beta}_n^{(0)} - \beta)}{s_L^2}, \quad (1.8)$$

where $s_L^2 = \frac{1}{m}(Y - X\tilde{\beta}_n^{(0)})^T(Y - X\tilde{\beta}_n^{(0)})$ and $m = n - p$.

Hence, the PTLE is defined by

$$\hat{\beta}_n^{PTL} = \tilde{\beta}_n^L - (\tilde{\beta}_n^L - \hat{\beta}_n^{RL})I(T_n^L \leq T_{n,\alpha}^L), \quad (1.9)$$

where $I(A)$ denotes the indicator of the set A and $T_{n,\alpha}^L$ is the upper α -level critical value of the exact distribution of T_n^L under \mathcal{H}_o .

Also, we consider the shrinkage LASSO estimator (SLE), analogous to the James and Stein (1961) rule, incorporating $\tilde{\beta}_n^L$, $\hat{\beta}_n^{RL}$ and the test statistic for \mathcal{H}_o in a smoother manner. We denote this by $\hat{\beta}_n^{SL}$ and defined by

$$\hat{\beta}_n^{SL} = \tilde{\beta}_n^L - k_n(\tilde{\beta}_n^L - \hat{\beta}_n^{RL})[T_n^L]^{-1}, \quad k_n = \frac{m(q-2)}{(m+2)}, \quad (1.10)$$

where k_n is the shrinkage constant.

The estimator $\hat{\beta}_n^{SLE}$ may go past the estimator $\hat{\beta}_n^{RL}$. Thus, we define the positive-rule shrinkage LASSO estimator (PRSLE) given by

$$\hat{\beta}_n^{PRSLE} = \hat{\beta}_n^{SL} - (1 - k_n[T_n^L]^{-1})I(T_n^L \leq k_n)(\tilde{\beta}_n^L - \hat{\beta}_n^{RL}). \quad (1.11)$$

We note that, as the test based on T_n^L is consistent against fixed β such that $H\beta \neq \mathbf{h}$, the PTLE, SLE and PRLE are asymptotically equivalent to the ULE for fixed alternative. Hence, we will investigate the asymptotic risks under local alternatives and compare the respective performance of the estimators.

2 Asymptotic Properties

We consider the following contiguous sequence of alternatives

$$K_{(n)} : H\beta = \mathbf{h} + n^{-1/2}\boldsymbol{\xi}, \quad \boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_q)^T.$$

Theorem 2.1. Under $K_{(n)}$ and regularity conditions (1.3) and (1.4) provided that $\lambda = O(\sqrt{n})$, as $n \rightarrow \infty$, distributions of T_n^L , ULE, RLE, PTLE, SLE and PRLE are given by

- (i) $\lim_{n \rightarrow \infty} P\left(\sqrt{n}(\tilde{\beta}_n^L - \beta) \leq \mathbf{x}\right) = \Phi_p(\mathbf{x}; \mathbf{0}, \sigma^2 C^{-1}),$
- (ii) $\lim_{n \rightarrow \infty} P\left(\sqrt{n}(\hat{\beta}_n^{RL} - \beta) \leq \mathbf{x}\right) = \Phi_p(\mathbf{x} + \boldsymbol{\delta}; \mathbf{0}, \sigma^2 A), \quad \boldsymbol{\delta} = C^{-1}H^T(HC^{-1}H^T)^{-1}\boldsymbol{\xi}$
and $A = C^{-1} - C^{-1}H^T(HC^{-1}H^T)^{-1}HC^{-1}.$
- (iii) $\lim_{n \rightarrow \infty} P(T_n^L \leq x) = H_q(x; \Delta^2); \quad \Delta^2 = \sigma^{-2}(\boldsymbol{\delta}^T C \boldsymbol{\delta}),$
- (iv) $\lim_{n \rightarrow \infty} P\left(\sqrt{n}(\hat{\beta}_n^{PTL} - \beta) \leq \mathbf{x}\right) = H_q(\chi_q^2(\alpha); \Delta^2)\Phi_p(\mathbf{x} + \boldsymbol{\delta}, \mathbf{0}, \sigma^2 A)$
 $+ \int_{E(\boldsymbol{\delta})} \Phi_p(\mathbf{x} + \mathbf{Z}, \mathbf{0}, \sigma^2(C^{-1} - A))d\Phi_p(\mathbf{Z}, \mathbf{0}, \sigma^2(HC^{-1}H^T)),$
where $E(\boldsymbol{\delta}) = \{\mathbf{Z} : (\mathbf{Z} + \boldsymbol{\delta})^T(HC^{-1}H^T)^{-1}(\mathbf{Z} + \boldsymbol{\delta}) \geq \chi_q^2(\alpha)\},$
- (v) $\sqrt{n}(\hat{\beta}_n^{SL} - \beta) \rightarrow_d \mathbf{Z} - k \left\{ \frac{\sigma^2 C^{-1}H^T(HC^{-1}H^T)^{-1}(H\mathbf{Z} + \boldsymbol{\xi})}{(H\mathbf{Z} + \boldsymbol{\xi})^T(HC^{-1}H^T)^{-1}(H\mathbf{Z} + \boldsymbol{\xi})} \right\},$
- (vi) $\sqrt{n}(\hat{\beta}_n^{PRL} - \beta) \rightarrow_d \mathbf{Z} - k \left\{ \frac{\sigma^2 C^{-1}H^T(HC^{-1}H^T)^{-1}(H\mathbf{Z} + \boldsymbol{\xi})}{(H\mathbf{Z} + \boldsymbol{\xi})^T(HC^{-1}H^T)^{-1}(H\mathbf{Z} + \boldsymbol{\xi})} \right\}$
 $- C^{-1}H^T(HC^{-1}H^T)^{-1}(H\mathbf{Z} + \boldsymbol{\xi}) \left\{ 1 - \frac{k\sigma^2}{(H\mathbf{Z} + \boldsymbol{\xi})^T(HC^{-1}H^T)^{-1}(H\mathbf{Z} + \boldsymbol{\xi})} \right\}$
 $\times I(\sigma^{-2}(H\mathbf{Z} + \boldsymbol{\xi})^T(HC^{-1}H^T)^{-1}(H\mathbf{Z} + \boldsymbol{\xi}) < k),$

where $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 C^{-1})$ and $k = q - 2$ and $\phi_p(\cdot; \boldsymbol{\mu}, \Sigma)$ to indicate the c.d.f. of the p -variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ , respectively. $H_\nu(\cdot; \Delta^2)$ denotes the c.d.f. of the χ^2 -distribution with degree of freedom ν and non centrality parameter $\Delta^2/2$.

Assume that, for an estimator $\hat{\beta}_n^*$ of β , $G^*(\mathbf{x}) = \lim_{n \rightarrow \infty} P\left(\sqrt{n}(\hat{\beta}_n^* - \beta) \leq \mathbf{x}\right).$

If the asymptotic c.d.f. exists, then the asymptotic distributional bias (ADB) and quadratic bias (ADQB) are given by

$$b(\hat{\beta}_n^*) = \lim_{n \rightarrow \infty} E \left[\sqrt{n}(\hat{\beta}_n^* - \beta) \right] = \int \mathbf{x} dG_p(\mathbf{x}),$$

and $B(\hat{\beta}_n^*) = \sigma^{-2}[b(\hat{\beta}_n^*)]^T C [b(\hat{\beta}_n^*)]$, respectively, where $\sigma^2 C^{-1}$ is the MSE-matrix of $\tilde{\beta}_n^L$ as $n \rightarrow \infty$ provided that $\lambda_n = O(\sqrt{n})$. Defining the asymptotic distributional MSE-matrix (ADMSE) as

$$M(\hat{\beta}_n^*) = \int \mathbf{x} \mathbf{x}^T dG_p(\mathbf{x}),$$

we have the asymptotic distributional quadratic risk (ADQR) of $\hat{\beta}_n^*$ given by

$$R(\hat{\beta}_n^*; W) = \text{tr} \left(W M(\hat{\beta}_n^*) \right).$$

Now, based on Corollary 1.2, we can easily obtain the ADB, ADQB, ADMSE and ADQR of the estimators.

Theorem 2.2. *The ADB, ADQB, ADMSE and ADQR of the estimators are given as*

- (i) $b_1(\tilde{\beta}_n^L) = \mathbf{0}, B_1(\tilde{\beta}_n^L) = 0, R_1(\tilde{\beta}_n^L; W) = \sigma^2 \text{tr}(W C^{-1}),$ and $M_1(\tilde{\beta}_n^L) = \sigma^2 C^{-1},$
- (ii) $b_2(\hat{\beta}_n^{RL}) = -\boldsymbol{\delta}, B_2(\hat{\beta}_n^{RL}) = \Delta^2, R_2(\hat{\beta}_n^{RL}; W) = \sigma^2 \text{tr}(W(C^{-1} - A)) + \boldsymbol{\delta}^T W \boldsymbol{\delta},$
and $M_2(\hat{\beta}_n^{RL}) = \sigma^2(C^{-1} - A) + \boldsymbol{\delta} \boldsymbol{\delta}^T,$
- (iii) $b_3(\hat{\beta}_n^{PTL}) = -\boldsymbol{\delta} H_{q+2}(\chi_q^2(\alpha); \Delta^2), B_3(\hat{\beta}_n^{PTL}) = \Delta^2 \{H_{q+2}(\chi_q^2(\alpha); \Delta^2)\}^2,$
 $R_3(\hat{\beta}_n^{PTL}; W) = \sigma^2 \text{tr}(W C^{-1}) - \sigma^2 \text{tr}(W(C^{-1} - A)) H_{q+2}(\chi_q^2(\alpha); \Delta^2) + \boldsymbol{\delta}^T W \boldsymbol{\delta} Z(\alpha; \Delta^2),$
 $M_3(\hat{\beta}_n^{PTL}) = \sigma^2 C^{-1} - \sigma^2(C^{-1} - A) H_{q+2}(\chi_q^2(\alpha); \Delta^2) + \boldsymbol{\delta} \boldsymbol{\delta}^T Z(\alpha; \Delta^2),$
- (iv) $b_4(\hat{\beta}_n^{SL}) = -k \boldsymbol{\delta} E[\chi_{q+2}^{-2}(\Delta^2)],$ where $k = \lim_{n \rightarrow \infty} k_n = q - 2, B_4(\hat{\beta}_n^{SL}) = k^2 \Delta^2 \{E[\chi_{q+2}^{-2}(\Delta^2)]\},$
 $R_4(\hat{\beta}_n^{SL}; W) = \sigma^2 \text{tr}(W C^{-1}) - k \sigma^2 \text{tr}(C^{-1} - A) X(\Delta^2) + k(k + 4) \boldsymbol{\delta}^T W \boldsymbol{\delta} E[\chi_{q+4}^{-4}(\Delta^2)],$
 $M_4(\hat{\beta}_n^{SL}) = \sigma^2 C^{-1} - k \sigma^2(C^{-1} - A) X(\Delta^2) + k(k + 4) \boldsymbol{\delta} \boldsymbol{\delta}^T E[\chi_{q+4}^{-4}(\Delta^2)],$
- (v) $b_5(\hat{\beta}_n^{PRL}) = b_4(\hat{\beta}_n^{SL}) - \boldsymbol{\delta} E[(1 - k \chi_{q+2}^{-2}(\Delta^2)) I(\chi_{q+2}^2(\Delta^2) \leq k)],$
 $B_5(\hat{\beta}_n^{PRL}) = \Delta^2 \{k E[\chi_{q+2}^{-2}(\Delta^2)] - E[(1 - k \chi_{q+2}^{-2}(\Delta^2)) I(\chi_{q+2}^2(\Delta^2) \leq k)]\}^2,$
 $R_5(\hat{\beta}_n^{PRL}; W) = R_4(\hat{\beta}_n^{SL}; W) - \sigma^2 \text{tr}(C^{-1} - A) E[(1 - k \chi_{q+2}^{-2}(\Delta^2))^2 I(\chi_{q+2}^2(\Delta^2) \leq k)]$
 $\quad - \boldsymbol{\delta}^T W \boldsymbol{\delta} Q(\Delta^2),$
 $M_5(\hat{\beta}_n^{PRL}) = M_4(\hat{\beta}_n^{SL}) - \sigma^2(C^{-1} - A) E[(1 - k \chi_{q+2}^{-2}(\Delta^2))^2 I(\chi_{q+2}^2(\Delta^2) \leq k)] - \boldsymbol{\delta} \boldsymbol{\delta}^T Q(\Delta^2).$

where

$$Z(\alpha; \Delta^2) = 2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2),$$

$$\begin{aligned} X(\Delta^2) &= 2E[\chi_{q+2}^{-2}(\Delta^2)] - kE[\chi_{q+4}^{-2}(\Delta^2)], \\ Q(\Delta^2) &= 2E[(1 - k\chi_{q+2}^{-2}(\Delta^2))I(\chi_{q+2}^2(\Delta^2) \leq k)] - E[(1 - k\chi_{q+4}^{-2}(\Delta^2))I(\chi_{q+4}^2(\Delta^2) \leq k)], \end{aligned}$$

and

$$\begin{aligned} E[\chi_{q+s}^{-2}(\Delta^2)] &= E_r(q + s - 2 + 2r)^{-1}, \\ E[\chi_{q+s}^{-4}(\Delta^2)] &= E_r[(q + s - 2 + 2r)(q + s - 4 + 2r)]^{-1} \\ E[\chi_{q+s}^{-2}(\Delta^2)I(\chi_{q+s}^2(\Delta^2) < k)] &= E_r(q + s - 2 + 2r)^{-1}H_{q+s-2+2r}(k; 0), \\ E[\chi_{q+s}^{-4}(\Delta^2)I(\chi_{q+s}^2(\Delta^2) < k)] &= E_r[(q + s - 2 + 2r)(q + s - 4 + 2r)]^{-1}H_{q+s-4+2r}(k; 0), \end{aligned}$$

where E_r stands for the expectation with respect to a Poisson variable r with mean $\Delta^2/2$.

Since the structures of derived asymptotic results in Theorem 2.2, are the same as those in Saleh (2006) and Saleh et al. (2014) under least square theory, we leave comparisons between the estimators to save the space.

3 Graphical Results

In this section, some graphical illustrations will be provided for asymptotic relative efficiencies (AREs) of the estimators. We assume $p = 5$, $q = 3$ and

$$\boldsymbol{\beta} = (3, 1, 0, -1, -2)^T, \quad H = \begin{bmatrix} 1 & 0 & -1 & 0 & -1 \\ 1 & 0 & -1 & -1 & -1 \\ 0 & -1 & -1 & 0 & 0 \end{bmatrix}, \quad \mathbf{h} = (0, 0, 0)^T \text{ and } \boldsymbol{\xi} = (0.05, 0.05, 0.05)^T.$$

In Figure 1, the ARE for any estimator $\hat{\boldsymbol{\beta}}_n^*$ is evaluated as

$$ARE = \frac{\text{Risk}(\tilde{\boldsymbol{\beta}}_n^L)}{\text{Risk}(\hat{\boldsymbol{\beta}}_n^*)}.$$

From Figure ??, it can be concluded that the ARE of the RLE relative to the LASSO is a decreasing function. The ARE of the PTLE relative to the LASSO is a decreasing function w.r.t. Δ^2 , and also for small level of significance α . The LASSO estimator performs better than the PTLE. Finally SLE performs better than the LASSO estimator.

References

- Arashi, M. and Nadarajah, S. (2014) On singular elliptical models, *Communication in Statistics: Theory and Methods*, Accepted.
- James, W. and Stein, C. (1961) Estimation with Quadratic Loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, 361–379.
- Knight K., and Fu, W. (2000). Asymptotic for lasso-type estimators, *Annals of Statistics*, **28**(5), 1356–1378.
- Saleh, A. K. M. E. (2006). *The Preliminary Test and Stein-Type Estimation with Applications*, John Wiley, London.
- Saleh, A. K. Md. E., Arashi, M. and Tabatabaey, S. M. M. (2014). *Statistical Inference for Models with Multivariate t -Distributed Errors*, John Wiley, New Jersey.
- Srivastava, M. S. (1971). On fixed width confidence bounds for regression parameters, *The Annals of Mathematical Statistics*, **42**, 1403–1411.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.

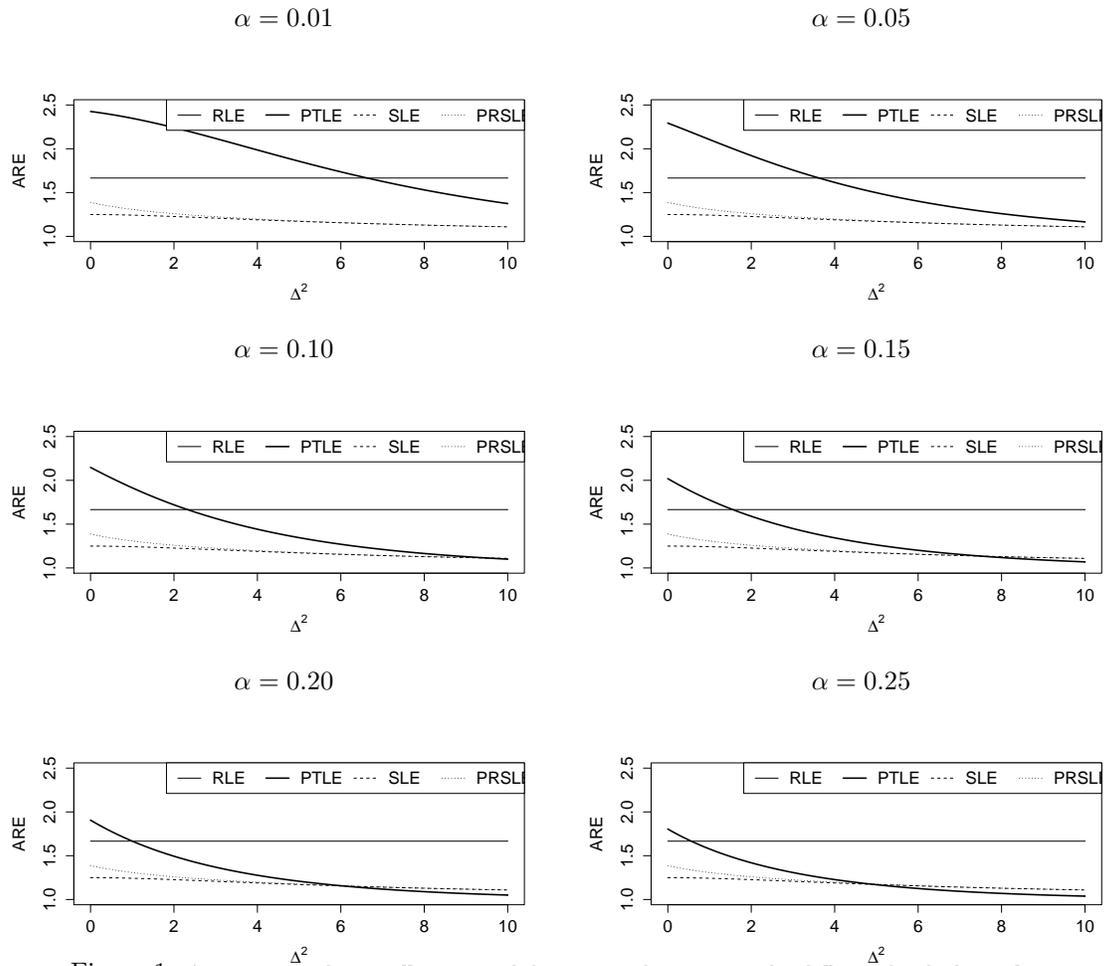


Figure 1: Asymptotic relative efficiencies of the proposed estimators for different level of significance.



Stochastic Differential Equations for SIS and SIR epidemic models

Mahmood Parsamanesh^{1,2}, Rahman Farnoosh²

¹Zabol, University of zabol, Faculty of sciences

²Tehran, Iran university of science and technology, Faculty of mathematical sciences

Abstract: Two epidemic models, Susceptible-Infectious-susceptible (SIS) and Susceptible-Infectious-Removed (SIR), are introduced in this article. These models are expressed in the form of system of ordinary differential equations, their dynamics are determined and finally the systems of stochastic differential equations are derived for them. moreover, numerical solutions of systems of ODEs and SDEs for epidemic models are compared by some examples.

Keywords Epidemic models, SIS model, SIR model, Initial reproduction number, Stochastic differential equations.

Mathematics Subject Classification (2010): 92D30 60H10 60H35.

1 Introduction

The emerging and spreading diseases have led to an interest in infectious diseases. Mathematical models have become important tools in analyzing the spread and control of infectious diseases [Hethcote, H. W. \(2000\)](#). Many models for the spread of infectious diseases in populations have been analyzed mathematically by ordinary differential equations and applied to specific diseases. But considering real life is not free of randomness, using stochastic models can be more real [Zhao and Jiang \(2013\)](#). SIS model is introduced and is formulated by ODEs in section 2. Then system of SDEs for this model is obtained in the next section. Deterministic and stochastic SIR models are the subjects of sections 4 and 5. We have numerical examples and computer simulations in the last section.

¹M.Parsamanesh : m.parsamanesh@uoz.ac.ir

2 Deterministic SIS Epidemic Model

In SIS epidemic model, individuals in the population are classified according to disease status, either susceptible or infectious. In this model individuals in susceptible class denoted by X_1 are transmitted to class X_2 , the infectious individuals, after becoming infectious. But after recovery, infected individuals return to the susceptible class. Suppose $x_1(t)$ and $x_2(t)$ are the number of individuals in each class in time t and the total population size in the beginning is N :

$$x_1(0) + x_2(0) = N$$

Diagram in Figure 1 illustrates the dynamics of the SIS epidemic model. Bold arrows show infection

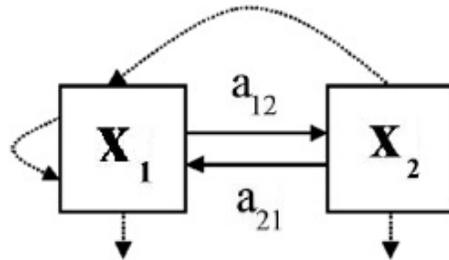


Figure 1: diagram of SIS epidemic model

(arrow from X_1 to X_2) and recovery (arrow from X_2 to X_1) and dotted arrows show birth and death. Suppose individuals in population are infected and recovered with rates β (contact rate) and γ (recovery rate), respectively. Moreover suppose every birth is accompanied by a death and so total population size of two classes is equal to constant N at every time:

$$x_1(t) + x_2(t) = N$$

Considering Figure 1 and letting $a_{12} = \beta \frac{x_2}{N}$ and $a_{21} = \gamma$ one can obtain the deterministic SIS model form as follows.

$$\begin{cases} \frac{dx_1}{dt} = -\beta \frac{x_2}{N} x_1 + \gamma x_2 \\ \frac{dx_2}{dt} = \beta \frac{x_2}{N} x_1 - \gamma x_2 \end{cases} \quad (2.1)$$

As mentioned earlier, initial condition of this problem is $:x_1(0) + x_2(0) = N$. The dynamics of model 2.1 are determined by the basic reproduction number which is the number of secondary infections

caused by one infected individual in an entirely susceptible population [Frase et al \(2009\)](#). For model 2.1, the basic reproduction number is defined as follows:

$$R_0 = \frac{\beta}{\gamma}$$

The asymptotic behaviours of model 2.1 are summarized in the following theorem [Braue and Castillo \(2001\)](#).

Theorem 2.1. *Let $x_1(t)$ and $x_2(t)$ be a solution to model 2.1.*

- (1) *If $R_0 \leq 1$, then $\lim(x_1(t), x_2(t)) = (N, 0)$ (disease-free equilibrium)*
- (2) *If $R_0 > 1$, then $\lim(x_1(t), x_2(t)) = (\frac{N}{R_0}, N(1 - \frac{1}{R_0}))$ (endemic equilibrium).*

3 Stochastic SIS epidemic model

Assume $t \in [0, \infty)$ and $x_1(t)$ and $x_2(t)$ are continuous variables in $[0, N]$. In this case $X = (x_1, x_2)^T$ and $\Delta X = (\Delta x_1, \Delta x_2)^T$ are two dimensional stochastic vectors. Moreover assume we choose time interval Δt so small such that only one change can occur between two populations X_1 and X_2 independently. Now consider the Figure 1 for SIS model. The possible changes in the two populations and their corresponding probability are introduced in Table 1.

Table 1: possible change in SIS model and their probability

Event	Probability	Change
from X_1 to X_2	$p_1 = a_{12}x_1\Delta t$	$\Delta X^{(1)} = (-1, 1)^T$
from X_2 to X_1	$p_2 = a_{21}x_2\Delta t$	$\Delta X^{(2)} = (1, -1)^T$
no change	$p_3 = 1 - (p_1 + p_2)$	$\Delta X^{(3)} = (0, 0)^T$

Using Table 1, the expected change and covariance matrix are determined for $\Delta X = (\Delta x_1, \Delta x_2)^T$ as follows:

$$E[\Delta X] = \sum_{i=1}^3 p_i \Delta X^{(i)} = \begin{pmatrix} -a_{12}x_1 + a_{21}x_2 \\ a_{12}x_1 - a_{21}x_2 \end{pmatrix} \Delta t, \tag{3.1}$$

$$E[\Delta X \cdot \Delta X^T] = \sum_{i=1}^3 p_i \Delta X^{(i)} \Delta X^{(i)T} = (a_{12}x_1 + a_{21}x_2) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \Delta t. \tag{3.2}$$

Stochastic differential equation of the following Ito form determines the dynamics of two populations X_1 and X_2 in the model. Keller (2011)

$$dX_t = \mu dt + B dW_t, \quad (3.3)$$

where $\mu = E[\Delta X]/\Delta t$ is drift coefficient, $B = V^{\frac{1}{2}}$ is diffusion coefficient if we let $V = E[\Delta X \cdot \Delta X^T]/\Delta t$, and $W_t = (w_1, w_2)^T$ is a two dimensional wiener process in SDE 3.3. One can obtain B by properties of the matrices as follow Higham N. J. (2008)

$$B = \sqrt{a_{12}x_1 + a_{21}x_2} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \quad (3.4)$$

Now, as the deterministic case if we let $a_{12} = \beta \frac{x_2}{N}$ and $a_{21} = \gamma$, from relations 3.1 and 3.4 the SDE 3.3 becomes the following system of SDEs

$$d \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} -\beta \frac{x_1 x_2}{N} + \gamma x_2 \\ \beta \frac{x_1 x_2}{N} - \gamma x_2 \end{pmatrix} dt + \sqrt{\beta \frac{x_1 x_2}{N} + \gamma x_2} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} d \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \quad (3.5)$$

It is sufficient to take the second row of the system 3.5 for number of infected individuals due to constant population. Moreover by properties of the gaussian distribution we obtain a SDE with one dimensional wiener process:

$$dx_2 = \left(\beta \frac{(N - x_2)x_2}{N} - \gamma x_2 \right) dt + \left(\sqrt{\beta \frac{(N - x_2)x_2}{N} + \gamma x_2} \right) dw. \quad (3.6)$$

4 Deterministic SIR epidemic model

In SIR model infected individuals doesn't return to susceptible class after recovery but they are removed and are transmitted to immune class. Therefore in this model population is classified to either susceptible, infectious or immune that are denoted by X_1 , X_2 and X_3 respectively. Assume the number of individuals in each of these classes is $x_1(t)$, $x_2(t)$ and $x_3(t)$, and total population size is constant:

$$x_1(t) + x_2(t) + x_3(t) = N$$

Figure 2 shows dynamics of this model. Therefore this model determined by following system of ODEs:

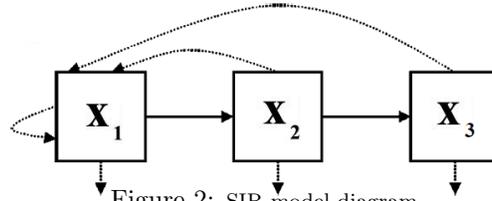


Figure 2: SIR model diagram

$$\begin{cases} \frac{dx_1}{dt} = -\beta \frac{x_2}{N} x_1 \\ \frac{dx_2}{dt} = \beta \frac{x_2}{N} x_1 - \gamma x_2 \\ \frac{dx_3}{dt} = \gamma x_2 \end{cases} \quad (4.1)$$

where β and γ are contact rate and recovery rate respectively. Asymptotic behavior of SIR model is summarized in the following theorem Braue and Castillo (2001).

Theorem 4.1. Let $x_1(t)$, $x_2(t)$ and $x_3(t)$ be a solution to model 4.1.

- (1) If $R_0 \leq 1$, then $\lim x_2(t) = 0$ (disease-free equilibrium)
- (2) If $R_0 > 1$, then $\lim (x_1(t), x_2(t), x_3(t)) = (\frac{N}{R_0}, 0, N(1 - \frac{1}{R_0}))$ (endemic equilibrium).
- (3) if $R_0 \frac{x_1(0)}{N} > 1$ then there is an initial increase in the number of infected cases $x_2(t)$ (epidemic), but if $R_0 \frac{x_1(0)}{N} \leq 1$ then $x_2(t)$ decreases monotonically to zero (disease-free equilibrium).

5 Stochastic SIR epidemic model

Assume $t \in [0, \infty)$ and $x_1(t)$, $x_2(t)$ and $x_3(t)$ are continuous random variables in $[0, N]$. As SIS model we consider $X = (x_1, x_2)^T$ and $\Delta X = (\Delta x_1, \Delta x_2)^T$ because of constant total population size and value of $x_3(t)$ can be determined by values of $x_1(t)$ and $x_2(t)$. moreover, possible changes between X_1 and X_2 and corresponding probabilities by choosing Δt enough small, are as in Table 2. The expected

Table 2: possible change in SIS model and their probability

Event	Probability	Change
from X_1 to X_2	$p_1 = \beta \frac{x_1 x_2}{N} \Delta t$	$\Delta X^{(1)} = (-1, 1)^T$
from X_2 to X_1	$p_2 = \gamma x_2 \Delta t$	$\Delta X^{(2)} = (1, -1)^T$
no change	$p_3 = 1 - (p_1 + p_2)$	$\Delta X^{(3)} = (0, 0)^T$

value and covariance matrix are as follow

$$E[\Delta X] = \begin{pmatrix} -\beta \frac{x_1 x_2}{N} \\ \beta \frac{x_1 x_2}{N} - \gamma x_2 \end{pmatrix} \Delta t, \tag{5.1}$$

$$E[\Delta X . \Delta X^T] = \beta \frac{x_1 x_2}{N} \begin{pmatrix} 1 & -1 \\ -1 & 1 + \frac{\gamma N}{\beta x_1} \end{pmatrix} \Delta t. \tag{5.2}$$

Therefore by calculating $\mu = E[\Delta X]/\Delta t$ and $B = V^{\frac{1}{2}}$ where $V = E[\Delta X . \Delta X^T]/\Delta t$, as drift and diffusion coefficients respectively, the fallowing system of SDEs can be obtained for SIR model

$$d \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} -\beta \frac{x_1 x_2}{N} \\ \beta \frac{x_1 x_2}{N} - \gamma x_2 \end{pmatrix} dt + \frac{\sqrt{\beta x_1 x_2 / N}}{\sqrt{2 + \frac{\gamma N}{\beta x_1} + 2\sqrt{\frac{\gamma N}{\beta x_1}}}} \begin{pmatrix} 1 + \frac{\gamma N}{\beta x_1} & -1 \\ -1 & 1 + \frac{\gamma N}{\beta x_1} + \sqrt{\frac{\gamma N}{\beta x_1}} \end{pmatrix} d \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{5.3}$$

6 Numerical examples

The general form of a d-dimensional Ito SDE is of the form

$$dX_t = a(t, x_t)dt + b(t, X_t)dW_t, \tag{6.1}$$

where $a = (a^1, a^2, \dots, a^d)^T$ is d-dimensional drift coefficient, $d \times m$ -dimensional matrix $b = (b^{ij})$ is diffusion coefficient and $W_t = (W_t^1, W_t^2, \dots, W_t^m)^T$ is m-dimensional wiener process. The system is named one dimensional if wiener process $\{W_t\}_{t_0 < t < T}$ is one dimensional and multidimensional otherwise. Notice that if $b = (b^{ij}) \equiv 0$ we have a system of ODEs.

Definition 6.1. A discrete approximation $\bar{X} = \{\bar{X}_0, \bar{X}_1, \dots, \bar{X}_N\}$ (based on a step size Δ) is said to converge strongly with order γ towards the solution $X = \{X_t\}$ of 6.1 in time T , if there exist constants $K > 0$ (not depending on Δ) and $\delta_0 > 0$ such that for all $\Delta \in (0, \delta_0)$

$$E[|X_T - \bar{X}_N|] \leq K \Delta^\gamma.$$

As an extension of Euler scheme for ordinary differential equations we have Euler-Maruyama scheme for stochastic differential equations. This scheme is of strong order $\gamma = 0.5$ and constant step size Δ and is as follows.

$$X_{n+1} = X_n + a\Delta + b\Delta W_n \tag{6.2}$$

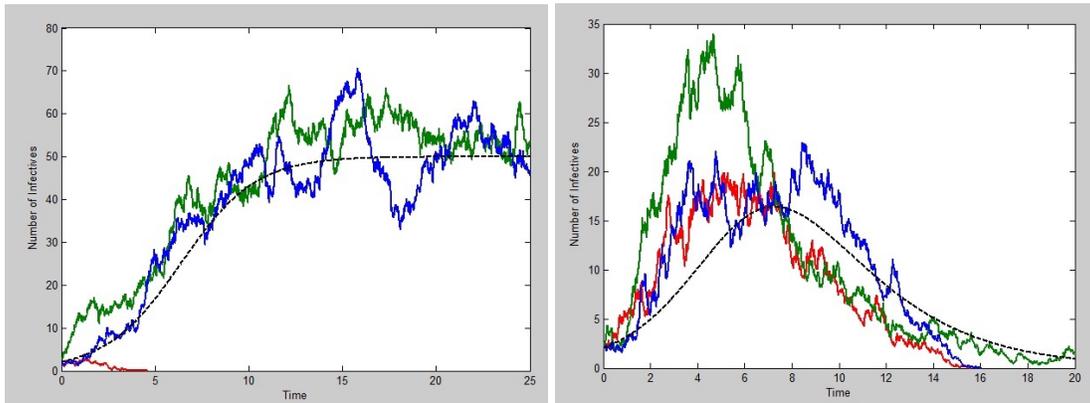


Figure 3: Three stochastic sample paths and deterministic solution for SIS model (left), and SIR model (right)

This scheme is applied for multidimensional wiener process as follows

$$X_{n+1}^k = X_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W^j, k = 1, \dots, d \quad (6.3)$$

where X_{n+1}^k is k -th component of solution vector X_{n+1} . This scheme and schemes of higher order can be found in Kloeden and Platen (1992).

Example 6.2. Consider SDE 3.6 for SIS model with parameters given by $N = 100, \beta = 1$ and $\gamma = 0.5$ and initial infected individuals $x_2(0) = 2$. For numerical solution assume $\Delta t = 0.01$ using Euler scheme for solving deterministic differential equation and Euler-Maruyama scheme for stochastic differential equation. Three sample paths of the stochastic SIS epidemic model accompanied by solution of deterministic model until time $T = 25$ are graphed in Figure 3 (left picture). The computations have been carried out with mathematical software MATLAB R2011a and codes as in Higham D. J. (2001). With these parameters we have $R_0 = 2$ and according to Theorem 2.1 the deterministic solution approaches the endemic equilibrium $\bar{x}_2 = 50$. Notice that sample path of stochastic model are continuous and one of them vanishes to zero and two other have same behavior as deterministic case.

Example 6.3. Consider system of SDEs 5.3 for stochastic SIR epidemic model with parameters same as previous example and initial vector $(x_1(0), x_2(0)) = (98, 2)$. With these parameters R_0 and $R_0 \frac{x_1(0)}{N}$ are both greater than one and so deterministic model must has an increase in beginning and then approaches to disease-free equilibrium sate with $\bar{x}_2 = 0$. Three sample paths of stochastic SIR model are compared with solution of deterministic model in Figure 3 (right picture). Epidemic is observable at beginning of all sample paths.

Conclusion

In this paper two prominent epidemic models, SIS model and SIR model, were considered and were expressed in system of ordinary differential equations and dynamic of them was given, also. Considering possible change and their probabilities and calculating expected vector and covariance matrix we got system of stochastic differential equations. Finally, behaviors of stochastic models were compared with deterministic model numerically.

References

- Brauer, F., Castillo-Chavez, C. (2001). *Mathematical Models in Population Biology and Epidemiology*, Springer, Berlin Heidelberg New York
- Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., ... & Roth, C. (2009). Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* **324** (5934), 1557-1561.
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *Society for Industrial and Applied Mathematics*, **42**(4), 599-653.
- Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, **43**(3), 525-546.
- Higham, N. J. (2008). *Functions of matrices: theory and computation*. Siam.
- Keller, A. A. (2011). Population biology models with time-delay in a noisy environment. *environment. parameters*, **1**, 2.
- Kloeden, P. E., Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin Heidelberg New York.
- Y. Zhao and D. Jiang. (2013). Dynamics of stochastically perturbed SIS epidemic model with vaccination, *Abstract and Applied Analysis*, **vol. 2013**, Hindawi Publishing Corporation.



An Approximation to Finite- and Infinite-time Ruin Probabilities of Compound Poisson Processes

Amir T. Payandeh Najafabadi

Department of Mathematical Sciences, Shahid Beheshti University, G.C. Evin,
1983963113, Tehran, Iran.

Abstract: Consider the problem of evaluating infinite-time (or finite-time) ruin probability under a given compound Poisson surplus process. Such problem leads to a integro-differential equation which cannot be solve analytically, in the most of situations. This article approximates claim size distribution by a finite mixture exponential, say Hyperexponential, distribution. Then, it restates the corresponding integro-differential equation as a solvable ordinary differential equation (or a partial differential equation for finite-time ruin probability). Application of our findings has been given though a simulation study.

Keywords Ruin probability, Compound Poisson Processes, mixture exponential (Hyperexponential) distribution, Heavy-tailed distributions.

Mathematics Subject Classification (2010):97K50, 97K60, 65L70.

1 Introduction

Consider the following compound Poisson process

$$U_t = u + ct - \sum_{j=1}^{N(t)} X_j, \quad (1.1)$$

where X_1, X_2, \dots are a sequence of i.i.d. random variables with common density function $f_X(\cdot)$, $N(t)$ is a Poisson process with intensity rate λ , u and c stand for initial wealth/reserve and premium of the process, respectively.

The finite-time and infinite-time ruin probabilities for the above compound Poisson process are,

respectively, denoted by $\psi(u; T)$ and $\psi(u)$ and, respectively, defined by $\psi(u; T) = P(\tau_u \leq T)$ and $\psi(u) = P(\tau_u < \infty)$, where τ_u is the hitting time, i.e., $\tau_u := \inf\{t : U_t \leq 0 | U_0 = u\}$.

Melnikov (2011, §3.2), among others, established that an infinite-time ruin probability $\psi(u)$ under a compound Poisson process can be restated as the following integro-differential equation

$$c\tilde{\psi}^{(1)}(u) - \lambda\tilde{\psi}(u) + \lambda \int_0^u \tilde{\psi}(u-x)f_X(x)dx = 0, \quad (1.2)$$

where $\tilde{\psi}(u) = 1 - \psi(u)$ and $\lim_{u \rightarrow \infty} \psi(u) = 0$.

Pervozvansky (1998) showed that a finite-time ruin probability $\psi(u; T)$ under a compound Poisson process can be restated as the following partial integro-differential equation

$$c \frac{\partial \tilde{\psi}(u; T)}{\partial u} - c \frac{\partial \tilde{\psi}(u; T)}{\partial T} - \lambda \tilde{\psi}(u; T) + \lambda \int_0^u \tilde{\psi}(u-x; T) f_X(x) dx = 0, \quad (1.3)$$

where $\tilde{\psi}(u; T) = 1 - \psi(u; T)$ $\lim_{u \rightarrow \infty} \psi(u; T) = 0$ for all $T > 0$ and $\psi(u; 0) = 0$ for all $u \geq 0$.

Since the compound Poisson surplus process plays a vital role in many actuarial models, several authors studies ruin probability under surplus process (1.1). An excellent review for infinite-time ruin probability can be found in Asmussen & Albrecher (2010). For finite-time ruin probability: Pervozvansky (1998) showed for exponential claim size distribution partial integro-differential equation (1.3) can be transformed into a second-order partial differential equation. Albrecher et al. (2001) considered a compound Poisson surplus process with constant force of real interest. Then, they restated finite-time ruin probability $\psi(u; T)$ as a gamma series expansion. Makroglou (2003) provided a global Lagrange type approximation in the z -space for $\psi(u; T)$ under surplus process (1.1). Avram et al. (2011) and Tran (2015) employed the Padé approximant method to approximate $\psi(u; T)$ under surplus process (1.1).

This article in the first step approximates claim size density function $f_X(\cdot)$ with a finite mixture exponential, say Hyperexponential, density function $f_X^*(\cdot)$. Then, it transforms two integro-differential equations (1.2) and (1.3), respectively, into an ordinary differential equation (ODE) and a partial differential equation (PDE). A simulation study has been conducted to show practical application of our findings.

The rest of this article is organized as follows. Some mathematical background for the problem has

been collected in Section 2. Section 3 provides the main contribution of this article. Applications of the results have been given in Section 4.

2 Preliminaries

From hereafter now, we set $\sum_{j=a}^b A_j = 0$, for $b < a$.

The following recalls the exponential type T functions which plays a vital role in the rest of this article. An $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ function f is said to be of exponential type T on \mathbb{C} if there are positive constants M and T such that $|f(\omega)| \leq M \exp\{T|\omega|\}$, for $\omega \in \mathbb{C}$. The Fourier transforms of exponential type functions are continuous functions which are infinitely differentiable everywhere and are given by a Taylor series expansion over every compact interval, see Champeney (1987, page 77) and Walnut (2002, page 81) for more details.

From the Hausdorff-Young Theorem, one can observe that if $\{s_n\}$ is a sequence of functions converging, in $L^2(\mathbb{R})$ sense, to s . Then, the Fourier transforms of s_n converge, in $L^2(\mathbb{R})$ sense, to the Fourier transform of s , see Pandey (1996) for more details. Using Lukacs (1987)'s method, Kucerovsky & Payandeh Najafabadi (2015) showed that most of the common distributions do have characteristic functions that can be extend to meromorphic functions.

The following from Merzon & Sadov (2011) and Sadov (2013) recalls Hausdorff-Young inequality for the Laplace transform. Suppose $h(\cdot)$ is a given and nonnegative function that $f \in L^1(\mathbb{R}^+) \cap L^2(\mathbb{R}^+)$. Then, $\|h\|_2 \leq \frac{1}{\sqrt{\pi}} \|\mathcal{L}(h)\|_2$, where $\mathcal{L}(h)$ stands for the Laplace transform. The Schwarz integrability condition states that in situation that all partial derivatives of a bivariate function exist and are continuous, one may change order of partial derivatives, see Belding & Mitchell (2008, Page 259) for more details.

The following lemma provides useful results for the next section. Suppose $k(\cdot)$ is a given and differentiable function and $y(\cdot)$ is an unknown function that satisfy

$$\int_0^x y(t) \left(\sum_{i=1}^n \omega_i \mu_i e^{-\mu_i(x-t)} \right) dt = k(x), \quad x \geq 0, \quad (2.1)$$

where ω_i , μ_i and μ_i are some given and nonnegative constants. Then, the above integral equation

function for a n -component Hyperexponential distribution is given by

$$f_X^*(x) = \sum_{i=1}^n \omega_i \mu_i e^{-\mu_i x}, \quad x \geq 0. \quad (2.2)$$

Feldmann & Whitt (1998) showed that, one may approximate a large class of distributions, including several heavy tail distributions such as Pareto and Weibull distributions, arbitrarily closely, by Hyperexponential distributions. Feller (1971, page 435) established that a survival function at x^γ , for all $x > 0$, is a completely monotone function if and only if its corresponding density function is a mixture of Weibull distributions with fixed shape parameter $1/\gamma$. Jewell (1982) showed that any Weibull distribution with shape parameter less than 1 can be restated as a Hyperexponential distributions.

Using the Hausdorff-Young Theorem, the following provides error bound for approximating the claim size density function $f_X(\cdot)$ by Hyperexponential density function $f_X^*(\cdot)$, given by (2.2). Suppose random claim size X is surplus process (1.1) has density function $f_X(\cdot)$ and characteristic function $\theta_X(\cdot)$. Moreover, suppose that characteristic function $\theta_X(\cdot)$ is (or can be extend to) a meromorphic function. Then, (1) density function of compound sum $S(t) = \sum_{i=1}^{N(t)} X_i$, say $f_{S(t)}(\cdot)$, can be approximated by density function $f_{S^*(t)}(\cdot)$, where $S(t) = \sum_{i=1}^{N(t)} Y_i$ and Y_i is a n -component Hyperexponential distribution; (2) Error bound for such approximation satisfies $\|f_{S(t)} - f_{S^*(t)}\|_2 \leq \lambda t e^{-\lambda t} \|\theta_X - \theta_Y\|_2$, where $\theta_Y(s) = \sum_{j=1}^n \omega_j \mu_j / (\mu_j + s\sqrt{-1})$. *Proof.* Using the Hausdorff-Young Theorem, one may conclude that $\|f_{S(t)} - f_{S^*(t)}\|_2 \leq \|e^{\lambda t(\theta_X - 1)} - e^{\lambda t(\theta_Y - 1)}\|_2$. The rest of proof arrives by using the fact that ψ_X and θ_Y are (or can be extend to) two meromorphic functions. \square

3 Ruin Probability

This section utilizes integro-differential Equations (1.2) and (1.3) to derive an approximate formula for the infinite (and finite)-time ruin probability of a compound Poisson process (1.1). We seek an analytical solution $\tilde{\psi}(\cdot)$ which is an exponential type function. In the other word, we assume: $|\tilde{\psi}(\omega)| \leq M e^{T|\omega|}$, $\omega \in \mathbb{C}$, for some real numbers M and T in \mathbb{R} . If this assumption is not met, as might be the case if, for example, there are point masses in $\psi(\cdot)$, our method works, but our error bounds may not be valid anymore.

The following theorem provides an $(n+1)$ -order ODE for infinite-time ruin probability $\psi(\cdot)$ in the

situation that claim size distribution X has been approximated by an n -component Hyperexponential density function $f_X^*(\cdot)$. Suppose claim size density function $f_X(\cdot)$ has been approximated by an n -component Hyperexponential density function $f_X^*(\cdot)$. Then, infinite-time survivals probability $\tilde{\psi}(\cdot)$ of a compound Poisson process (1.1) can be approximated by infinite-time survivals probability $\tilde{\psi}_*(\cdot)$ which can be evaluated using the following $(n + 1)$ -order ODE.

$$\begin{aligned}
 0 = & \sum_{i=1}^n \lambda \omega_i \mu_i \tilde{\psi}_*^{(n-1)}(u) + \sum_{i=1}^n \sum_{i \neq j}^n \lambda \omega_i \mu_i \mu_j \tilde{\psi}_*^{(n-2)}(u) - \sum_{i=1}^n \sum_{i \neq j}^n \sum_{k > j, k \neq i}^n \lambda \omega_i \mu_i \mu_j \mu_k \tilde{\psi}_*^{(n-3)}(u) \\
 & + \sum_{i=1}^n \sum_{i \neq j}^n \sum_{k > j, k \neq i}^n \sum_{l > k, l \neq i}^n \lambda \omega_i \mu_i \mu_j \mu_k \mu_l \tilde{\psi}_*^{(n-4)}(u) + \dots + (-1)^n \sum_{i=1}^n \prod_{j \neq i}^n \mu_j \tilde{\psi}_*^{(0)}(u) \\
 & - \left[\lambda \tilde{\psi}_*^{(n)}(u) - c \tilde{\psi}_*^{(n+1)}(u) \right] - \sum_{i=1}^n \mu_i \left[\lambda \tilde{\psi}_*^{(n-1)}(u) - c \tilde{\psi}_*^{(n)}(u) \right] \\
 & - \sum_{i=1}^n \sum_{j \neq i}^n \mu_i \mu_j \left[\lambda \tilde{\psi}_*^{(n-2)}(u) - c \tilde{\psi}_*^{(n-1)}(u) \right] \\
 & - \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k > j, k \neq i}^n \mu_i \mu_j \mu_k \left[\lambda \tilde{\psi}_*^{(n-3)}(u) - c \tilde{\psi}_*^{(n-2)}(u) \right] - \dots - \prod_{i=1}^n \mu_i \left[\lambda \tilde{\psi}_*^{(0)}(u) - c \tilde{\psi}_*^{(1)}(u) \right],
 \end{aligned}$$

with boundary conditions that satisfy $c \tilde{\psi}_*^{(m)}(0) - \lambda \tilde{\psi}_*^{(m-1)}(0) + \lambda \sum_{j=0}^{m-2} \tilde{\psi}_*^{(j)}(0) f^{(m-2-j)}(0) = 0$, for $m = 1, \dots, n$. *Proof.* An application of Lemma (2) by changing $k(u) \mapsto -c \tilde{\psi}_*^{(1)}(u) + \lambda \tilde{\psi}_*(u)$, $y(u) \mapsto \tilde{\psi}_*(u)$, and $\omega_i \mapsto \lambda \omega_i$ lead to the desired result. \square

Using the fact that $\tilde{\psi}_*(0) = 1 - \lambda E(X)/c$, (Kaas et al., 2008, Page 104) the above boundary condition equation leads to: $\tilde{\psi}_*^{(1)}(0) = \tilde{\psi}_*(0) \frac{\lambda}{c}$,

$$\begin{aligned}
 \tilde{\psi}_*^{(2)}(0) &= \tilde{\psi}_*(0) \left[\left(\frac{\lambda}{c}\right)^2 - \left(\frac{\lambda}{c}\right) f_X(0) \right], \\
 \tilde{\psi}_*^{(3)}(0) &= \tilde{\psi}_*(0) \left[\left(\frac{\lambda}{c}\right)^3 - 2 f_X(0) \left(\frac{\lambda}{c}\right)^2 - \left(\frac{\lambda}{c}\right) f_X^{(1)}(0) \right], \\
 \tilde{\psi}_*^{(4)}(0) &= \tilde{\psi}_*(0) \left[\left(\frac{\lambda}{c}\right)^4 - 3 f_X(0) \left(\frac{\lambda}{c}\right)^3 + \left(\frac{\lambda}{c}\right)^2 [-2 f_X^{(1)}(0) + f_X^2(0)] - \left(\frac{\lambda}{c}\right) f_X^{(2)}(0) \right], \\
 \tilde{\psi}_*^{(5)}(0) &= \tilde{\psi}_*(0) \left[\left(\frac{\lambda}{c}\right)^5 - 4 \left(\frac{\lambda}{c}\right)^4 f_X(0) + \left(\frac{\lambda}{c}\right)^3 [-3 f_X^{(1)}(0) + 3 f_X^2(0)] + \left(\frac{\lambda}{c}\right)^2 [-2 f_X^{(2)}(0) + 2 f_X(0) f_X^{(1)}(0)] - \left(\frac{\lambda}{c}\right) f_X^{(3)}(0) \right], \\
 \tilde{\psi}_*^{(6)}(0) &= \tilde{\psi}_*(0) \left[\left(\frac{\lambda}{c}\right)^6 - 5 \left(\frac{\lambda}{c}\right)^5 f_X(0) + \left(\frac{\lambda}{c}\right)^4 [6 f_X^2(0) - 4 f_X^{(1)}(0)] + \left(\frac{\lambda}{c}\right)^3 [-3 f_X^{(2)}(0) + 6 f_X(0) f_X^{(1)}(0) - f_X^3(0)] \right. \\
 & \left. + \tilde{\psi}_*(0) \left[\left(\frac{\lambda}{c}\right)^2 [2 f_X^{(3)}(0) + 2 f_X(0) f_X^{(2)}(0) + f_X^{(1)}(0) f_X^{(1)}(0)] - \left(\frac{\lambda}{c}\right) f_X^{(4)}(0) \right] \right], \\
 \tilde{\psi}_*^{(7)}(0) &= \tilde{\psi}_*(0) \left[\left(\frac{\lambda}{c}\right)^7 - 6 \left(\frac{\lambda}{c}\right)^6 f_X(0) + \left(\frac{\lambda}{c}\right)^5 [10 f_X^2(0) - 5 f_X^{(1)}(0)] + \left(\frac{\lambda}{c}\right)^4 [-4 f_X^{(2)}(0) + 7 f_X(0) f_X^{(1)}(0) - 4 f_X^{(3)}(0)] \right. \\
 & \left. + \tilde{\psi}_*(0) \left(\frac{\lambda}{c}\right)^3 [2 f_X^{(3)}(0) + 2 f_X(0) f_X^{(1)}(0) + 3 f_X^{(1)}(0) f_X^{(1)}(0) - f_X^{(3)}(0) + 4 f_X(0) f_X^{(2)}(0) - 3 f_X^2(0) f_X^{(1)}(0)] \right. \\
 & \left. + \tilde{\psi}_*(0) \left[\left(\frac{\lambda}{c}\right)^2 [-2 f_X^{(4)}(0) + f_X(0) f_X^{(3)}(0) + 2 f_X^{(1)}(0) f_X^{(2)}(0) + f_X^{(3)}(0)] - \left(\frac{\lambda}{c}\right) f_X^{(5)}(0) \right] \right] \text{ and so on.}
 \end{aligned}$$

The following provides error bound for approximating infinite-time survivals probability $\tilde{\psi}(\cdot)$ by $\tilde{\psi}_*(\cdot)$. Suppose claim size density function $f_X(\cdot)$ has been approximated by an n -component Hyperexponential density function $f_X^*(\cdot)$. Then, the infinite-time survival probability $\tilde{\psi}(u)$ of compound Poisson process (1.1) can be approximated by $\tilde{\psi}_*(u)$, given by Theorem (3), and its error satisfies $\|\psi(u) - \psi_*(u)\|_2 \leq \frac{c\lambda\tilde{\psi}(0)}{\sqrt{\pi a_1^2}} \left\| \varphi_X(s) - \sum_{j=1}^n \frac{\omega_j \mu_j}{\mu_j + s} \right\|_2$, where $a_1 = \sup\{\varphi_X(s), \sum_{j=1}^n \frac{\omega_j \mu_j}{\mu_j + s}\}$ and $\varphi_X(s)$ stands for the characteristic function of random claim X . *Proof.* Application of the Hausdorff-Young for Laplace transform (Lemma 2) along with fact that $\mathcal{L}(g'(x); x; s) = s\mathcal{L}(g(x); x; s) - g(0)$ and $\mathcal{L}(\int_0^x (g(x-y)f(y)dy; x; s) = \mathcal{L}(g(x); x; s)\mathcal{L}(f(x); x; s)$, one may conclude that

$$\|\psi(u) - \psi_*(u)\|_2 \leq \frac{1}{\sqrt{\pi}} \|\mathcal{L}(\tilde{\psi}) - \mathcal{L}(\tilde{\psi}_*)\|_2 = \frac{1}{\sqrt{\pi}} \left\| \frac{c\tilde{\psi}(0)}{cu - \lambda + \lambda\mathcal{L}(f)} - \frac{c\tilde{\psi}(0)}{cu - \lambda + \lambda\mathcal{L}(f^*)} \right\|_2.$$

Application of inequality $\|1/h_1 - 1/h_2\|_2 \leq a^{-2}\|h_1 - h_2\|_2$, where $a = \sup\{h_1, h_2\}$, from Kucerovsky & Payandeh Najafabadi (2009) completes the desired proof. \square

The following theorem provides an $(n + 1)$ -order PDE for finite-time ruin probability $\psi(\cdot)$ in the situation that claim size distribution X has been approximated by an n -component Hyperexponential distribution function. Suppose claim size density function $f_X(\cdot)$ has been approximated by an n -component Hyperexponential density function $f_X^*(\cdot)$. Then, finite-time survivals probability $\tilde{\psi}(u; T)$ of a compound Poisson process (1.1) can be approximated by finite-time survivals probability $\tilde{\psi}_*(u; T)$ which can be evaluated using the following $(n + 1)$ -order PDE. $0 = \sum_{i=1}^n \lambda \omega_i \mu_i \frac{\partial^{n-1}}{\partial u^{n-1}} \tilde{\psi}_*(u; T) + \sum_{i=1}^n \sum_{i \neq j} \lambda \omega_i \mu_i \mu_j \frac{\partial^{n-2}}{\partial u^{n-2}} \tilde{\psi}_*(u; T) - \sum_{i=1}^n \sum_{i \neq j} \sum_{k > j, \neq i} \lambda \omega_i \mu_i \mu_j \mu_k \frac{\partial^{n-3}}{\partial u^{n-3}} \tilde{\psi}_*(u; T) + \sum_{i=1}^n \sum_{i \neq j} \sum_{k > j, \neq i} \sum_{l > k, \neq i} \lambda \omega_i \mu_i \mu_j \mu_k \mu_l \frac{\partial^{n-4}}{\partial u^{n-4}} \tilde{\psi}_*(u; T) - \dots + (-1)^n \prod_{j \neq i} \mu_j \frac{\partial^0}{\partial u^0} \tilde{\psi}_*(u; T) - \left[\lambda \frac{\partial^n}{\partial u^n} \tilde{\psi}_*(u; T) - c \frac{\partial^{n+1}}{\partial u^{n+1}} \tilde{\psi}_*(u; T) + c \frac{\partial^{n+1}}{\partial T \partial u^n} \tilde{\psi}_*(u; T) \right] - \sum_{i=1}^n \mu_i \left[\lambda \frac{\partial^{n-1}}{\partial u^{n-1}} \tilde{\psi}_*(u; T) - c \frac{\partial^n}{\partial u^n} \tilde{\psi}_*(u; T) + c \frac{\partial^n}{\partial T \partial u^{n-1}} \tilde{\psi}_*(u; T) \right] - \sum_{i=1}^n \sum_{j \neq i} \mu_j \left[\lambda \frac{\partial^{n-2}}{\partial u^{n-2}} \tilde{\psi}_*(u; T) - c \frac{\partial^{n-1}}{\partial u^{n-1}} \tilde{\psi}_*(u; T) + c \frac{\partial^{n-1}}{\partial T \partial u^{n-2}} \tilde{\psi}_*(u; T) \right] - \sum_{i=1}^n \sum_{j \neq i} \sum_{k > j, \neq i} \mu_i \mu_j \mu_k \left[\lambda \frac{\partial^{n-3}}{\partial u^{n-3}} \tilde{\psi}_*(u; T) - c \frac{\partial^{n-2}}{\partial u^{n-2}} \tilde{\psi}_*(u; T) + c \frac{\partial^{n-2}}{\partial T \partial u^{n-3}} \tilde{\psi}_*(u; T) \right] - \dots - \prod_{i=1}^n \mu_i \left[\lambda \frac{\partial^0}{\partial u^0} \tilde{\psi}_*(u; T) - c \frac{\partial^1}{\partial u^1} \tilde{\psi}_*(u; T) + c \frac{\partial^1}{\partial T \partial u^0} \tilde{\psi}_*(u; T) \right]$, and boundary conditions that satisfy $c\tilde{\psi}_*^{(m)}(0; T) - c \frac{\partial}{\partial T} \tilde{\psi}_*^{(m-1)}(0; T) - \lambda \tilde{\psi}_*^{(m-1)}(0; T) + \lambda \sum_{j=0}^{m-2} \tilde{\psi}_*^{(j)}(0; T) f_X^{(n-2-j)}(0) = 0$, for $m = 1, \dots, n$, where $\tilde{\psi}_*^{(n)}(0; T) = \lim_{u \rightarrow 0} \frac{\partial^n}{\partial u^n} \tilde{\psi}_*(u; T)$. *Proof.* Using partial integro-differential equation (1.3) and the Schwarz integrability condition, one may change order of differentiation and obtain the above recursive

formula for boundary conditions. An application of Lemma (2) by changing $k(u) \mapsto -c \frac{\partial}{\partial u} \tilde{\psi}_*(u; T) + c \frac{\partial}{\partial T} \tilde{\psi}_*(u; T) + \lambda \tilde{\psi}_*(\cdot)$, $y(u) \mapsto \tilde{\psi}_*(u; T)$, and $\omega_i \mapsto \lambda \omega_i$ lead to the desired result. \square

Using the fact that $\tilde{\psi}_*(u; 0) = 1$, $\tilde{\psi}_*(0; T) = \int_0^{cT} F_{S,T}(x) dx / (cT)$, and $F_{S,T}(x) = P(\sum_{j=1}^{N(T)} X_j \leq x)$, for all $x \in \mathbb{R}^+$ (Asmussen & Albrecher, 2010, Page 121). One may compute the following from boundary conditions from recursive formula given by Theorem (3). $\tilde{\psi}_*^{(1)}(0; T) = \frac{\lambda}{c} \tilde{\psi}_*(0; T) + \frac{\partial}{\partial T} \tilde{\psi}_*(0; T)$,

$$\begin{aligned} \tilde{\psi}_*^{(2)}(0; T) &= \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^2 - \left(\frac{\lambda}{c}\right) f_X(0) \right] + \frac{\partial}{\partial T} \tilde{\psi}_*^{(1)}(0; T), \\ \tilde{\psi}_*^{(3)}(0; T) &= \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^3 - 2f_X(0) \left(\frac{\lambda}{c}\right)^2 - \left(\frac{\lambda}{c}\right) f_X^{(1)}(0) \right] + \frac{\partial}{\partial T} \tilde{\psi}_*^{(2)}(0; T), \\ \tilde{\psi}_*^{(4)}(0; T) &= \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^4 - 3f_X(0) \left(\frac{\lambda}{c}\right)^3 + \left(\frac{\lambda}{c}\right)^2 [-2f_X^{(1)}(0) + f_X^2(0)] - \left(\frac{\lambda}{c}\right) f_X^{(2)}(0) \right] + \frac{\partial}{\partial T} \tilde{\psi}_*^{(3)}(0; T), \\ \tilde{\psi}_*^{(5)}(0; T) &= \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^5 - 4\left(\frac{\lambda}{c}\right)^4 f_X(0) + \left(\frac{\lambda}{c}\right)^3 [-3f_X^{(1)}(0) + 3f_X^2(0)] + \left(\frac{\lambda}{c}\right)^2 [-2f_X^{(2)}(0) + 2f_X(0) f_X^{(1)}(0)] - \left(\frac{\lambda}{c}\right) f_X^{(3)}(0) \right] \\ &+ \frac{\partial}{\partial T} \tilde{\psi}_*^{(4)}(0; T), \\ \tilde{\psi}_*^{(6)}(0; T) &= \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^6 - 5\left(\frac{\lambda}{c}\right)^5 f_X(0) + \left(\frac{\lambda}{c}\right)^4 [6f_X^2(0) - 4f_X^{(1)}(0)] + \left(\frac{\lambda}{c}\right)^3 [-3f_X^{(2)}(0) + 6f_X(0) f_X^{(1)}(0) - f_X^3(0)] \right] \\ &+ \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^2 [2f_X^{(3)}(0) + 2f_X(0) f_X^{(2)}(0) + f_X^{(1)}(0) f_X^{(1)}(0)] - \left(\frac{\lambda}{c}\right) f_X^{(4)}(0) \right] + \frac{\partial}{\partial T} \tilde{\psi}_*^{(5)}(0; T), \\ \tilde{\psi}_*^{(7)}(0; T) &= \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^7 - 6\left(\frac{\lambda}{c}\right)^6 f_X(0) + \left(\frac{\lambda}{c}\right)^5 [10f_X^2(0) - 5f_X^{(1)}(0)] + \left(\frac{\lambda}{c}\right)^4 [10f_X(0) f_X^{(1)}(0) - 3f_X^3(0) - 4f_X^{(2)}(0) f_X(0)] \right] \\ &+ \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^3 [2f_X^{(3)}(0) + 2f_X(0) f_X^{(1)}(0) + 3f_X^{(1)}(0) f_X^{(1)}(0) - f_X^{(3)}(0) + 4f_X(0) f_X^{(2)}(0) - 3f_X^2(0) f_X^{(1)}(0)] \right] \\ &+ \tilde{\psi}_*(0; T) \left[\left(\frac{\lambda}{c}\right)^2 [-2f_X^{(4)}(0) + f_X(0) f_X^{(3)}(0) + 2f_X^{(1)}(0) f_X^{(2)}(0) + f_X^{(3)}(0)] - \left(\frac{\lambda}{c}\right) f_X^{(5)}(0) \right] + \frac{\partial}{\partial T} \tilde{\psi}_*^{(6)}(0; T), \text{ where } \tilde{\psi}_*^{(n)}(0; T) = \lim_{u \rightarrow 0} \frac{\partial^n}{\partial u^n} \tilde{\psi}_*(u; T). \end{aligned}$$

Using the central limit theorem for compound sum $\sum_{i=1}^{N(t)} X_i$ (see Kass et al., 2008, §2.5, or Schmidli 2006, §1.9), one may provide the following approximation for expression $\tilde{\psi}_*(0; T) = \int_0^{cT} F_{S,T}(x) dx / (cT)$

$$\tilde{\psi}_*(0; T) \approx \frac{1}{cT} \int_0^{cT} \Phi \left(\frac{x - \lambda T m_1}{\sqrt{\lambda T m_2}} \right) dx,$$

where $m_i = E(X^i)$, for $i = 1, 2$, and $\Phi(\cdot)$ stands for cumulative distribution function for standard normal distribution, see Kass et al (2008, §2.5) ,or Schmidli (2006, §1.9), for other parametric approximation approaches and Qin & Pitts (2012) for a nonparametric approximation approach. For heavy tailed random claim size X that the ordinary central limit theorem does not work properly. One has to employ an appropriated version of the central limit theorem, see Rachev (2003) and Bilik (2008), among others, for more details.

The following provides error bound for approximating finite-time survivals probability $\tilde{\psi}(u; T)$ by $\tilde{\psi}_*(u; T)$. Suppose claim size density function $f_X(\cdot)$ has been approximated by an n -component Hyperexponential density function $f_X^*(\cdot)$. Then, the infinite-time survival probability $\tilde{\psi}(u; T)$ of compound

Poisson process (1.1) can be approximated by $\tilde{\psi}_*(u; T)$, given by Theorem (3), and its error satisfies

$$\|\psi(u; T) - \psi_*(u; T)\|_2 \leq \frac{\lambda}{\sqrt{\pi}} \left[-\frac{c\tilde{\psi}(0; T)}{a_1^2} + \frac{a_2 T}{c} - c\tilde{\psi}(0; T)a_1^2 a_2^2 a_3 \right] \left\| \varphi_X(s) - \sum_{j=1}^n \frac{\omega_j \mu_j}{\mu_j + s} \right\|_2,$$

where $a_1 = \sup\{\varphi_X(s), \sum_{j=1}^n \frac{\omega_j \mu_j}{\mu_j + s}\}$, $a_2 = \sup\{1/s - c/A(s), 1/s - c/A_*(s)\}$, $a_3 = \sup\{e^{A(s)/cT}, e^{A_*(s)/cT}\}$, $A(s) = cs - \lambda + \lambda\varphi_X(s)$, $A_*(s) = cs - \lambda + \lambda \sum_{j=1}^n \omega_j \mu_j / (\mu_j + s)$, and $\varphi_X(s)$ stands for the characteristic function of random claim X . *Proof.* Taking the Laplace transform from both sides of Equation (1.3) leads to the following first-order PDE $A(s)\mathcal{L}(\tilde{\psi}(u; T); u; s) - c\tilde{\psi}(0; T) - c\frac{\partial}{\partial T}\mathcal{L}(\tilde{\psi}(u; T); u; s) = 0$, where $\mathcal{L}(\tilde{\psi}(u; 1); u; s) = 1/s$. Therefore, the Laplace transform of finite-time ruin probability for compound Poisson process (1.1) is $\mathcal{L}(\tilde{\psi}(u; T); u; s) = \frac{c\tilde{\psi}(0; T)}{A(s)} + \left(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A(s)}\right) e^{A(s)/cT}$. The above finding along with an application of the Hausdorff-Young for Laplace transform (Lemma 2) lead to

$$\begin{aligned} \|\psi(u; T) - \psi_*(u; T)\|_2 &\leq \frac{1}{\sqrt{\pi}} \|\mathcal{L}(\tilde{\psi}(u; T); u; s) - \mathcal{L}(\tilde{\psi}_*(u; T); u; s)\|_2 \\ &= \frac{1}{\sqrt{\pi}} \left\| \frac{c\tilde{\psi}(0; T)}{A(s)} + \left(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A(s)}\right) e^{A(s)/cT} - \frac{c\tilde{\psi}(0; T)}{A_*(s)} - \left(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A_*(s)}\right) e^{A_*(s)/cT} \right\|_2 \\ &\leq \frac{c\tilde{\psi}(0; T)}{\sqrt{\pi}a_1^2} \|A(s) - A_*(s)\|_2 + \frac{b}{\sqrt{\pi}} \left\| \frac{TA(s)}{c} - \frac{TA_*(s)}{c} \right\|_2 \\ &\quad + \frac{b}{\sqrt{\pi}} \left\| \ln\left(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A(s)}\right) - \ln\left(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A_*(s)}\right) \right\|_2, \end{aligned}$$

where the second inequality arrives by application of inequality $\|1/h_1 - 1/h_2\|_2 \leq a^{-2} \|h_1 - h_2\|_2$, where $a = \sup\{h_1, h_2\}$, from Kucerovsky & Payandeh Najafabadi (2009), triangle inequality, and the Mean value theorem (i.e., $(\exp\{A(s)/cT + \ln(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A(s)})\}) - \exp\{A_*(s)/cT + \ln(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A_*(s)})\}) / (A(s)/cT + \ln(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A(s)}) - A_*(s)/cT - \ln(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A_*(s)})) \leq b$ where $b = \sup\{A(s)/cT + \ln(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A(s)}), A_*(s)/cT + \ln(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A_*(s)})\}$). Application of inequality $\|\ln h_1 - \ln h_2\|_2 \leq \|h_1 - h_2\|_2/a$, where $a = \sup\{h_1, h_2\}$, from Kucerovsky & Payandeh Najafabadi (2009) completes the desired proof. \square

4 Simulation Study

Consider compound Poisson process (1.1) with intensity rate $\lambda = 1$ and premium $c = 1.1$. This section conducts two simulation studies to show practical application the about findings. Suppose random claim X in compound Poisson process (1.1) has been distributed according to Weibull(0.3,9,26053). Feldmann & Whitt (1998) using a three-moment matching algorithm showed that density function of random claim X can be approximated by the following 2-component Hyperexponential density

function $f_X^*(x) = 0.000095e^{-0.019x} + 1.348225e^{-1.355x}$. For infinite-time ruin probability: Application of Theorem (3) leads to the following second order ODE $1.1\tilde{\psi}_*^{(3)}(u)+0.5114\tilde{\psi}_*^{(2)}(u)+0.0026395\tilde{\psi}_*^{(1)}(u) = 0$ with initial conditions $\tilde{\psi}_*(0) = 0.0909090909$, $\lim_{u \rightarrow 0} \tilde{\psi}_*^{(1)}(u) = 0.08264462809$, and $\lim_{u \rightarrow 0} \tilde{\psi}_*^{(2)}(u) = -0.03629992491$.

Solving the above ODE, one may approximate finite-time survival probability $\tilde{\psi}(u)$ of compound Poisson process (1.1) by $\tilde{\psi}_*(u) = 0.9974815963 - 0.2101123939e^{-0.01502369720u} - 0.2873692025e^{-0.8589763028u}$.

Figure 1: Part (a) and Part (b) illustrate behavior for such approximated infinite-time ruin probability and its corresponding squared error, respectively. Such error has been evaluated by substitution approximated infinite-time ruin probability in integro-differential equation (1.2).

For finite-time ruin probability: Application of Theorem (3) leads to the following PDE

$$0 = 1.1 \frac{\partial^3}{\partial u^3} \tilde{\psi}_*(u; T) + 0.5114 \frac{\partial^2}{\partial u^2} \tilde{\psi}_*(u; T) - 0.0539995 \frac{\partial}{\partial u} \tilde{\psi}_*(u; T) - 1.1 \frac{\partial^3}{\partial u^2 \partial T} \tilde{\psi}_*(u; T) - 1.5114 \frac{\partial^2}{\partial u \partial T} \tilde{\psi}_*(u; T) - 0.0283195 \frac{\partial}{\partial T} \tilde{\psi}_*(u; T)$$

with initial conditions $\tilde{\psi}_*(u, 0) = 0$, $\tilde{\psi}_*(0; T) = \beta(T)$, $\lim_{u \rightarrow 20} \tilde{\psi}_*(u; T) = 1$, $\lim_{u \rightarrow 0} \frac{\partial}{\partial u} \tilde{\psi}_*(u; T) = 0.9091\beta(T) + \frac{\partial}{\partial T} \beta(T)$, where $\beta(T) = \frac{1}{1.1T} \int_0^{1.1T} \Phi\left(\frac{x-T}{\sqrt{29.36T}}\right) dx$.

Solving the above PDE, one may approximate finite-time survival probability $\tilde{\psi}(u; T)$ of compound Poisson process (1.1) by $\tilde{\psi}_*(u; T)$, that its behavior (for $T = 50, 100, 200$) has been illustrated in Part (c) of Figure 2. Part (d) of Figure 2 illustrates squared error of our approximation (for $T = 50, 100, 200$).

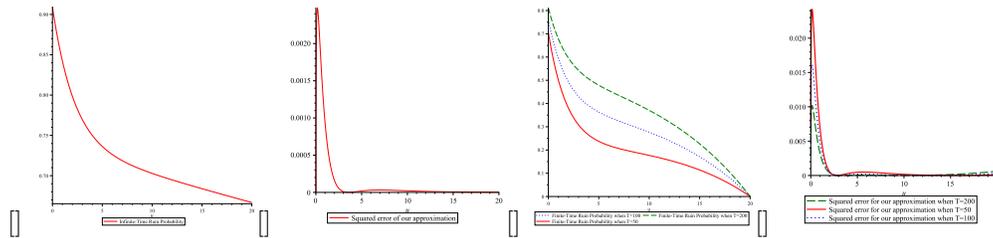


Figure 1: Part (a): Behavior for approximated infinite-time ruin probability $\psi_*(u)$; Part (b): Squared error for approximated infinite-time ruin probability $\psi_*(u)$; Part (c): Behavior for approximated finite-time ruin probability $\psi_*(u; T)$, for $T = 50, 100, 200$; and Part (d): Squared error for approximated finite-time ruin probability, for $T = 50, 100, 200$.

As one may observe, our squared error is less than 0.0025 and 0.025, for approximating infinite-time and finite-time ruin probability, respectively. Such error maybe reduced by increasing number of component in Hyperexponential density function. Suppose random claim X in compound Poisson process (1.1)

has been distributed according to Gamma(0.7310,1). Tran (2015) using the Padé approximant method showed that density function of random claim X can be approximated by the following 3-component Hyperexponential density function $f_X^*(x) = 0.8099e^{-3.2398x} + 0.3616e^{-1.4465x} + 0.5198e^{-1.0396x}$. For infinite-time ruin probability: Application of Theorem (3) leads to the following second order ODE $1.1\tilde{\psi}_*^{(4)}(u) + 5.29849\tilde{\psi}_*^{(3)}(u) + 6.479507012\tilde{\psi}_*^{(2)}(u) + 1.797919457\tilde{\psi}_*^{(1)}(u) = 0$ with initial conditions $\tilde{\psi}_*(0) = 0.3354861821$, $\lim_{u \rightarrow 0} \tilde{\psi}_*^{(1)}(u) = 0.3049874383$, $\lim_{u \rightarrow 0} \tilde{\psi}_*^{(2)}(u) = -0.2385743202$, and $\lim_{u \rightarrow 20} \tilde{\psi}_*(u) = 1$.

Solving the above ODE, one may approximate infinite-time survival probability $\tilde{\psi}(u)$ of compound Poisson process (1.1) by $\tilde{\psi}_*(u) = 1 - 0.013037e^{-3.073097u} - 0.008568e^{-1.349630u} - 0.642908e^{-0.394082u}$.

Figure 2: Part (a) and Part (b) illustrate behavior for such approximated infinite-time ruin probability and its corresponding squared error, respectively. Such error has been evaluated by substitution approximated infinite-time ruin probability in integro-differential equation (1.2).

For finite-time ruin probability: Application of Theorem (3) leads to the following PDE

$$0 = 1.1 \frac{\partial^4}{\partial u^4} \tilde{\psi}_*(u; T) + 5.29849 \frac{\partial^3}{\partial u^3} \tilde{\psi}_*(u; T) + 6.479507012 \frac{\partial^2}{\partial u^2} \tilde{\psi}_*(u; T) + 1.797919457 \frac{\partial}{\partial u} \tilde{\psi}_*(u; T) - 5.359146 \frac{\partial}{\partial T} \tilde{\psi}_*(u; T) - 10.5141 \frac{\partial^2}{\partial u \partial T} \tilde{\psi}_*(u; T) - 6.2985 \frac{\partial^3}{\partial u^2 \partial T} \tilde{\psi}_*(u; T) - 1.1 \frac{\partial^4}{\partial u^3 \partial T} \tilde{\psi}_*(u; T)$$

with initial conditions $\tilde{\psi}_*(u, 0) = 0$, $\tilde{\psi}_*(0; T) = \beta(T)$, $\lim_{u \rightarrow 20} \tilde{\psi}_*(u; T) = 1$, $\lim_{u \rightarrow 0} \frac{\partial}{\partial u} \tilde{\psi}_*(u; T) = 0.9091\beta(T) + \frac{\partial}{\partial T} \beta(T)$, $\lim_{u \rightarrow 0} \frac{\partial}{\partial u} \tilde{\psi}_*(u; T) = -0.71113\beta(T) + 0.9091 \frac{\partial}{\partial T} \beta(T) + \frac{\partial^2}{\partial T^2} \beta(T)$, where $\beta(T) = \frac{1}{1.1T} \int_0^{1.1T} \Phi\left(\frac{x-0.7309651999T}{\sqrt{0.7309651995T}}\right) dx$.

Solving the above PDE, one may approximate finite-time survival probability $\tilde{\psi}(u; T)$ of compound Poisson process (1.1) by $\tilde{\psi}_*(u; T)$, that its behavior (for $T = 50, 100, 200$) has been illustrated in Part (c) of Figure 2. Part (d) of Figure 2 illustrates squared error of our approximation (for $T = 50, 100, 200$).

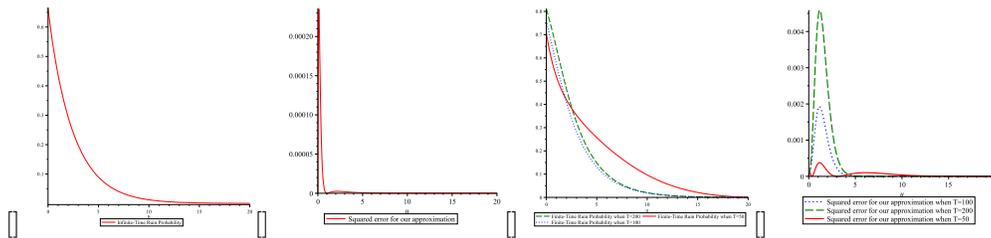


Figure 2: Part (a): Behavior for approximated infinite-time ruin probability $\psi_*(u)$; Part (b): Squared error for approximated infinite-time ruin probability $\psi_*(u)$; Part (c): Behavior for approximated finite-time ruin probability $\psi_*(u; T)$, for $T = 50, 100, 200$; and Part (d): Squared error for approximated finite-time ruin probability, for $T = 50, 100, 200$.

As one may observe, our squared error is less than 0.00025 and 0.005, for approximating infinite-time and finite-time ruin probability, respectively. Such error maybe reduced by increasing number of component in Hyperexponential density function.

It worthwhile to mention that: A given density function (or a density function corresponding to a given data set) can be approximated by a Hyperexponential distribution using a Matlab package called “bayesf”, see Sylvia (2008) for more details.

5 Conclusion and Suggestions

This article approximates claim size density function $f_X(\cdot)$ by a n -component Hyperexponential density function $f_X^*(\cdot)$. Then, it restates the problem of finding an infinite-time (or finite-time) ruin probability as a $(n + 1)$ -order ordinary differential equation (or a partial differential equation for finite-time ruin probability). Application of our findings has been given though a simulation study.

Certainly the following generalized Hyperexponential distribution can be provided a more accurate approximation in the situation that the true density function (or recorded data) has more than one mode. $g_X^{GHE}(x) = \sum_{i=1}^n \omega_i \mu_i e^{-\mu_i(x-b_i)} I_{[b_i, \infty)}(x)$. In such situation the finite and infinite ruin probabilities can be evaluated using the following lemma. Suppose claim size density function $f_X(\cdot)$ has been approximated by generalized Hyperexponential distribution $g_X^{GHE}(\cdot)$. The survival probability can be found by the following two inverse Laplace transforms.

(i) The infinite-time survival probability can be found by the following inverse Laplace transform

$$\tilde{\psi}(u) = \mathcal{L}^{-1} \left(\frac{c\tilde{\psi}(0)}{cs - \lambda + \lambda \sum_{i=1}^k \frac{\omega_i \mu_i}{\mu_i + s} e^{-sb_i}}; s; u \right)$$

(ii) The finite-time survival probability can be found by the following inverse Laplace transform

$$\tilde{\psi}(u; T) = \mathcal{L}^{-1} \left(\frac{c\tilde{\psi}(0; T)}{A^{**}(s)} + \left(\frac{1}{s} - \frac{c\tilde{\psi}(0; T)}{A^{**}(s)} \right) e^{A^{**}(s)/cT}; s; u \right), \text{ where } A^{**}(s) = cs - \lambda + \lambda \sum_{i=1}^k \frac{\omega_i \mu_i}{\mu_i + s} e^{-sb_i}.$$

Proof. The desired result arrives by taking a Laplace transform from both sides of equations (1.2) and (1.3) and solving corresponding first-order PDE with boundary condition $\tilde{\psi}(u; 0) = 1$ or $\mathcal{L}(\tilde{\psi}(u; 0); u; s) = 1/s$. \square

References

Albrecher, H., Teugels, J. L., & Tichy, R. F. (2001). On a gamma series expansion for the time-dependent probability of collective, ruin. *Insurance: Mathematics and Economics*, **29**(3), 345–355.

- Avram, F., Chedom, D. F., & Horváth, A. (2011). On moments based Padé approximations of ruin probabilities. *Journal of computational and applied mathematics*, **235**(10), 3215–3228.
- Asmussen, S., & Albrecher, H. (2010). *Ruin probabilities* (Vol. 14). World Scientific, New Jersey.
- Belding, D. F., & Mitchell, K. J. (2008). *Foundations of analysis* (Vol. 79). Courier Corporation, New York.
- Bilik, A. (2008). *Heavy-tail central limit theorem*. Lecture notes from University of California, San Diego.
- Champeney, D. C. (1987). *A handbook of Fourier theorems*. Cambridge University press, New York.
- Feldmann, A., & Whitt, W. (1998). Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance evaluation*, **31**(3-4), 245–279.
- Feller, W. (1971). *An introduction to probability and its applications*, Vol. II. Wiley, New York.
- Jewell, N. P. (1982). Mixtures of Exponential Distributions. *The Annals of Statistics*, **10**(2), 479–484.
- Kaas, R., Goovaerts, M., Dhaene, J., & Denuit, M. (2008). *Modern actuarial risk theory: using R* (Vol. 128). Springer-Verlag, Berlin Heidelberg.
- Kucerovsky, D., & Payandeh Najafabadi, A. T. (2009). An approximation for a subclass of the Riemann-Hilbert problems. *IMA journal of applied mathematics*, **74**(4), 533–547.
- Kucerovsky, D. Z. & Payandeh Najafabadi, A. T. (2015). Solving an integral equation arising from the Ruin probability of long-term Bonus–Malus systems. *Preprinted*.
- Lukacs, E. (1987). *Characteristic Functions*. Oxford University Press, New York.
- Makroglou, A. (2003). Integral equations and actuarial risk management: some models and numerics. *Mathematical Modelling and Analysis*, **8**(2), 143–154.
- Melnikov, A. (2011). *Risk analysis in finance and insurance*. CRC Press.
- Merzon, A., & Sadov, S. (2011). Hausdorff-Young type theorems for the Laplace transform restricted to a ray or to a curve in the complex plane. *arXiv preprint arXiv:1109.6085*.
- Pandey, J. (1996). *The Hilbert transform of Schwartz distributions and application*. John Wiley & Sons, New York.

- Pervozvansky, A. A. (1998). Equation for survival probability in a finite time interval in case of non-zero real interest force. *Insurance: Mathematics and Economics*, **23**(3), 287–295.
- Qin, L., & Pitts, S. M. (2012). Nonparametric estimation of the finite-time survival probability with zero initial capital in the classical risk model. *Methodology and Computing in Applied Probability*, **14**(4), 919–936.
- Rachev, S. T. (Ed.). (2003). *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance* (Vol. 1). Elsevier, Netherlands.
- Sadov, S. Y. (2013). Characterization of Carleson measures by the Hausdorff-Young property. *Mathematical Notes*, **94**(3-4), 551–558.
- Schmidli, H. (2006). *Lecture notes on Risk Theory*. Institute of Mathematics University of Cologne, German.
- Sylvia, F. S. (2008). *Finite mixture and Markov switching models: Implementation in MATLAB using the package bayesf Version 2.0*. Springer-Verlag, New York.
- Tran, D. X. (2015). Padé approximants for finite time ruin probabilities. *Journal of Computational and Applied Mathematics*, **278**, 130–137.
- Walnut, D. F. (2002). *An introduction to wavelet analysis*. 2nd ed. Birkhäuser publisher, New York.



Slashed generalized normal distribution: Model and properties

Atefe Pourchitsaz^{1,2}, Eisa Mahmoudi² and Hamide Jafari³

^{1,2,3}Department of Statistics, Yazd University, 89175-741, Yazd, IRAN

Abstract: In this paper a new generalized version of the normal distribution, called the slashed generalized normal distribution, is introduced. Probability distribution function of the new distribution is obtained. Also, moments, moment generating function and some other properties of the proposed distribution are derived.

Keywords Generalized normal distribution, Moments, Normal distribution, Slash distribution.

Mathematics Subject Classification (2010): 60E05; 62H10; 62H12.

1 Introduction

A distribution closely related to the normal distribution is the slash distribution. It can be represented as the quotient between to independent random variables, a normal one (numerator) and the power of uniform $(0, 1)$ distribution (denominator). Hence, we can say that a random variable S has a slash distribution if it can be represented as

$$S = \frac{Z}{U^{1/q}},$$

where $Z \sim N(0, 1)$, independent of $U \sim U(0, 1)$ and $q > 0$. In particular, as $q \rightarrow \infty$, the standard normal distribution follows. On the other hand, if $q = 1$, we obtain the canonic (standard) slash distribution, with density function given by

$$p(x) = \begin{cases} \frac{\phi(0) - \phi(x)}{x^2} & x \neq 0 \\ \frac{1}{2}\phi(0) & x = 0, \end{cases} \quad (1.1)$$

where ϕ represents the density function of the standard normal distribution (Johnson et al. (1995)). Rogers and Tukey (1972) and Mosteller and Tukey (1977) discuss about properties of this family. Kafadar (1982) studied maximum likelihood estimators for the location scale case. A multivariate version and an asymmetric multivariate version are developed by Wang and Genton (2006). Gomez et al. (2007) and Gomez and Venegas (2008) introduced the slash-elliptical family. Arslan (2008),

²Atefe Pourchitsaz: a.chitsaz@stu.yazd.ac.ir

studied asymmetric versions of this family. A symmetric extension of the multivariate slash distribution is studied by Arslan and Genc (2009) and a symmetric univariate generalization of the slash distribution is studied by Genc (2007). The maximum likelihood estimation for the parameters of the skew slash distribution is discussed by Arslan (2009). Gomez et al. (2009) used the slash-elliptical family to extend the Birnbaum-Saunders distribution. Olmos et al. (2012) introduced the slashed half-normal distribution.

A random variable X is said to have the generalized normal distribution with parameters μ , σ and s if its density function is given by

$$f(x) = K \exp\left\{-\left|\frac{x-\mu}{\sigma}\right|^s\right\}, \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0, s > 0,$$

where $K = \frac{s}{2\sigma\Gamma(1/s)}$. We denote this as $X \sim GN(\mu, \sigma, s)$. Nadarajah (2005) introduced this distribution as the generalization of normal and Laplace distributions where $s = 2$ and $s = 1$, respectively. The generalized normal distribution is bell-shaped and unimodal with mode at $x = \mu$.

The cdf of X can be written as

$$F(x) = \begin{cases} \frac{\Gamma(1/s, ((\mu-x)/\sigma)^s)}{2\Gamma(1/s)} & \text{if } x \leq \mu \\ 1 - \frac{\Gamma(1/s, ((x-\mu)/\sigma)^s)}{2\Gamma(1/s)} & \text{if } x > \mu, \end{cases}$$

where incomplete gamma function defined by

$$\Gamma(a, x) = \int_x^\infty t^{a-1} \exp(-t) dt.$$

Proposition 1.1. *Some properties of the generalized normal distribution are*

1. $E[X] = \mu$,
2. $Var(X) = \frac{\sigma^2\Gamma(3/s)}{\Gamma(1/s)}$,
3. $E[X^n] = \frac{\mu^n \sum_{k=0}^n \binom{n}{k} (\sigma/\mu)^k \{1 + (-1)^k\} \Gamma((k+1)/s)}{2\Gamma(1/s)}$,
4. $Skewness(X) = 0$,
5. $Kurtosis(X) = \frac{\Gamma(1/s)\Gamma(5/s)}{\Gamma^2(3/s)}$.

In this paper we introduce a new distribution that contains normal, Laplace and generalized normal distribution. This new distribution can be represented as the quotient between two independent random variables, a generalized normal one as numerator, and the power of uniform $(0, 1)$ distribution as denominator. Moments and some properties of proposed distribution are derived.

2 Slashed generalized normal distribution

In this section we give the stochastic representation and the density function of the slashed generalized normal distribution, and study some of its distributional properties.

We consider that the random variable Y has a slashed generalized normal distribution with parameters μ , σ , s and q if it can be represented as the ratio

$$Y = \frac{X}{U^{1/q}}, \quad (2.1)$$

where $X \sim GN(\mu, \sigma, s)$ and $U \sim U(0, 1)$ are independent, $-\infty < \mu < \infty$, $\sigma > 0$, $s > 0$ and $q > 0$. We denote this as $Y \sim SLGN(\mu, \sigma, s, q)$.

A distribution that plays an important role in this paper is the gamma distribution with cdf given by

$$G(x; \alpha, \beta) = \int_0^x g(t; \alpha, \beta) dt,$$

where $x > 0$, $\alpha > 0$ and $\beta > 0$ and $g(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$. The following theorem presents the pdf for the slashed generalized normal distribution that can be generated by representation (2.1).

Theorem 2.1. *Let $Y \sim SLGN(\mu, \sigma, s, q)$. Then, the density function of Y is given by*

$$f_Y(y; \mu, \sigma, s, q) = \begin{cases} \frac{K}{s} q \sigma^{q+1} \Gamma\left(\frac{q+1}{s}\right) |y - \mu|^{-(q+1)} G(|y - \mu|^s, \frac{q+1}{s}, \frac{1}{\sigma^s}) & \text{if } y \neq \mu \\ K \frac{q}{q+1} & \text{if } y = \mu, \end{cases}$$

where $K = \frac{s}{2\sigma\Gamma(1/s)}$, $\sigma > 0$, $s > 0$, $q > 0$ and G is the cdf of the gamma distribution.

Proof. Without loss of generality, we may suppose that $\mu = 0$. We define the transformation $Z = Y - \mu$, thus $Z \sim SLGN(0, \sigma, s, q)$.

Here we only give the proof for $z > 0$. The proof for $z < 0$ is similar and for $z = 0$, we know that the pdf of random variable Z is not continuous in zero but its discontinuity is removable type. Thus we substitute $\lim_{z \rightarrow 0} f_Z(z)$ for $z = 0$. Using the Equation (2.1), after obtaining its jacobian we have

$$f_Z(z) = K \frac{q}{z^{q+1}} \int_0^z t^q \exp\left\{-\left(\frac{t}{\sigma}\right)^s\right\} dt,$$

and making the variable transformation $u = t^s$ we have that

$$f_Z(z) = \frac{K}{s} \frac{q}{z^{q+1}} \int_0^{z^s} u^{\frac{q+1}{s}-1} \exp\left\{-\frac{u}{\sigma^s}\right\} du,$$

and the result follows after identifying the gamma density within the integral sign. When $z = 0$,

$$\begin{aligned} f_Z(0) &= \lim_{z \rightarrow 0^+} \frac{K}{s} q \sigma^{q+1} \Gamma\left(\frac{q+1}{s}\right) z^{-(q+1)} G\left(z^s, \frac{q+1}{s}, \frac{1}{\sigma^s}\right) \\ &= \lim_{z \rightarrow 0^-} \frac{K}{s} q \sigma^{q+1} \Gamma\left(\frac{q+1}{s}\right) (-z)^{-(q+1)} G\left((-z)^s, \frac{q+1}{s}, \frac{1}{\sigma^s}\right) \\ &= K \frac{q}{q+1} \end{aligned}$$

□

Remark 2.2. As $\mu = 0$, $\sigma = \sqrt{2}$, $s = 2$ and $q = 1$ we obtain the canonic slash distribution. Hence, the density function of Y is given by (1.1).

Figure 1 plots some possible shapes of the slashed generalized normal distribution for selected values of s and q by fixing $\mu = 0$ and $\sigma = 1$. The slashed generalized normal distribution is symmetric for all values.

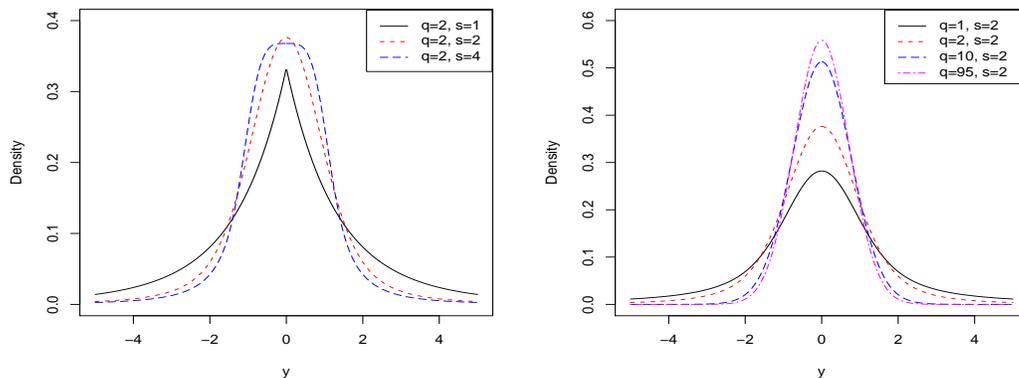


Figure 1: Plots of slashed generalized normal distribution for $\mu = 0$ and $\sigma = 1$.

Theorem 2.3. Let $Y|U = u \sim GN(0, u^{-1/q}\sigma, s)$ and $U \sim U(0, 1)$ then $Y \sim SGN(0, \sigma, s, q)$.

Proof. The marginal distribution of random variable Y can be written as

$$f_Y(y; 0, \sigma, s, q) = \int_0^1 f_{Y|U}(y|u) f_U(u) du = \int_0^1 \frac{su^{1/q}}{2\sigma\Gamma(1/s)} \exp\left\{-\left|\frac{yu^{1/q}}{\sigma}\right|^s\right\} du$$

and hence result follows making the variable transformation $t = yu^{1/q}$.

□

3 Moments of SLGN distribution

In this section, we obtain moments, skewness and kurtosis of random variable of slashed generalized normal distribution.

Theorem 3.1. *Let $Z \sim SLGN(0, 1, s, q)$. Then for $r = 1, 2, \dots$ it follows that the r th moment is given by*

$$E[Z^r] = \begin{cases} \frac{q}{q-r} \frac{1+(-1)^r}{2\Gamma(1/s)} \Gamma\left(\frac{r+1}{s}\right) & q > r \\ \text{dose not exist} & q \leq r. \end{cases}$$

Corollary 3.2. *Let $Y \sim SLGN(\mu, \sigma, s, q)$. Then for $n = 1, 2, \dots$ it follows that the n th moments of Y is given by*

$$\mu_n = E[Y^n] = \begin{cases} \frac{q\mu^n}{2\Gamma(1/s)} \sum_{r=0}^n \binom{n}{r} (\sigma/\mu)^r \frac{1+(-1)^r}{q-r} \Gamma\left(\frac{r+1}{s}\right) & q > n \\ \text{dose not exist} & q \leq n. \end{cases}$$

In particular, the four moments can be worked out as

$$\begin{aligned} E[Y] &= \mu, \\ E[Y^2] &= \mu^2 + \frac{q\sigma^4\Gamma(3/s)}{(q-2)\Gamma(1/s)}, \\ E[Y^3] &= \mu^3 + \frac{3q\mu\sigma^4\Gamma(3/s)}{(q-2)\Gamma(1/s)}, \\ E[Y^4] &= \mu^4 + \frac{6q\mu^2\sigma^4\Gamma(3/s)}{(q-2)\Gamma(1/s)} + \frac{q\sigma^8\Gamma(5/s)}{(q-4)\Gamma(1/s)}. \end{aligned}$$

The n th central moment can be obtained as

$$E[(Y - \mu)^n] = \frac{q}{q-n} \sigma^n \frac{1+(-1)^n}{2\Gamma(1/s)} \Gamma\left(\frac{n+1}{s}\right).$$

In particular, the first three central moments, skewness and kurtosis of Y are given by

$$\begin{aligned} Var(Y) &= \frac{q\sigma^2\Gamma(3/s)}{(q-2)\Gamma(1/s)}, \\ E[(Y - E(Y))^3] &= 0, \\ E[(Y - E(Y))^4] &= \frac{q\sigma^4\Gamma(5/s)}{(q-4)\Gamma(1/s)}, \\ Skewness(Y) &= 0, \\ Kurtosis(Y) &= \frac{(q-2)^2\Gamma(1/s)\Gamma(5/s)}{q(q-4)\Gamma^2(3/s)}. \end{aligned}$$

Figure 2 shows that the kurtosis measure depend on the parameters of s and q .

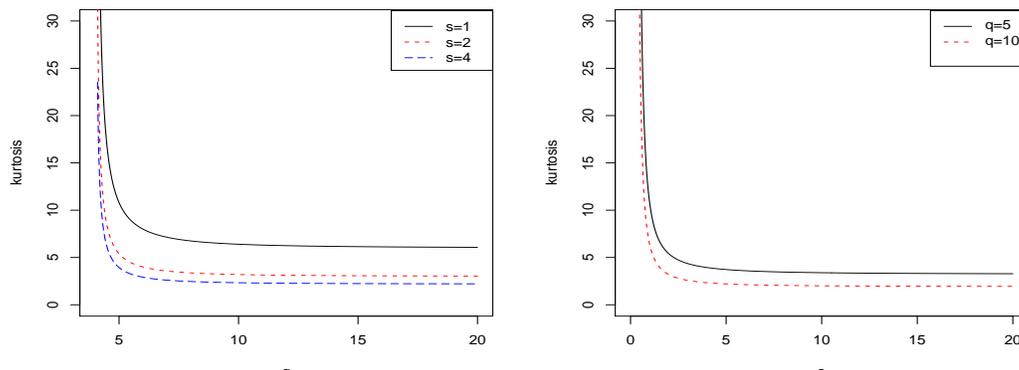


Figure 2: Plots of SLGN kurtosis for some values s and q .

Theorem 3.3. Let $Y \sim SLGN(\mu, \sigma, s, q)$, then the moment generating function for the random variable Y is given by

$$M_Y(t) = \frac{q}{2\Gamma(1/s)} \sum_{n=0}^{\infty} \sum_{k=0}^n \binom{n}{k} \frac{t^n}{n!} \mu^{n-k} \sigma^k \frac{1 + (-1)^k}{q - k} \Gamma\left(\frac{k+1}{s}\right).$$

Theorem 3.4. If $Y \sim \text{Gamma}(\frac{1}{s}, 1)$, then $V = \frac{W}{U^{1/q}}$ has $SLGN(\mu, \sigma, s, q)$ distribution, where $W = \mu + \sigma TY^{1/s}$ and T is an independent random variable of Y and taking values of 1 and -1 with equal probability.

The following algorithm gives method of generating random sample of size n from $SLGN(\mu, \sigma, s, q)$ distribution.

This algorithm has the following steps:

- Step 1: generating a sample of size n from $Y \sim \text{Gamma}(\frac{1}{s}, 1)$.
- Step 2: generating a random sample of size n from T using a random sample from $U(0, 1)$.
- Step 3: compute $W = \mu + \sigma TY^{1/s}$ for known values μ, σ and s .
- Step 4: calculate n dimensional vector $V = \frac{W}{U^{1/q}}$, where U is a random of size n from $U(0, 1)$ distribution.

Conclusion

In this paper we introduce a new distribution, called as slashed generalized normal distribution, that contains normal, Laplace and generalized normal distribution. This new distribution can be represented as the quotient between two independent random variables, a generalized normal one as numerator, and the power of uniform $(0, 1)$ distribution as denominator.

References

- Arslan, O. (2008). An alternative multivariate skew-slash distribution, *Statistics and Probability Letters*, **78**(16), 2756–2761.
- Arslan, O. (2009). Maximum likelihood parameter estimation for the multivariate skew-slash distribution, *Statistics and Probability Letters*, **79**, 2158–2165.
- Arslan, O. and Genc, A.I. (2009). A generalization of the multivariate slash distribution, *Journal of Statistical Planning and Inference*, **139**(3), 1164–1170.
- Genc, A.I. (2007). A generalization of the univariate slash by a scale-mixture exponential power distribution, *Communications in Statistics-Simulation and Computation*, **36**(5), 937–947.
- Gomez, H.W., Quintana, F.A. and Torres, F.J. (2007). A new family of slash-distributions with elliptical contours, *Statistics and Probability Letters*, **77**(7), 717–725.
- Gomez, H.W. and Venegas, O. (2008). Erratum to: A new family of slash-distributions with elliptical contours [*Statistics and Probability Letters*, **77**(2007), 717–725], *Statistics and Probability Letters*, **78**(14), 2273–2274.
- Gomez, H.W., Olivares-Pacheco, J.F. and Bolfarine, H. (2009). An extension of the generalized Birnbaum-Saunders distribution, *Statistics and Probability Letters*, **79**(3), 331–338.
- Johanson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous univariate distributions, Vol 1, 2nd edition*, John Wiley, New York.
- Kafadar, K. (1982). A biweight approach to the one-sample problem, *Journal of American Statistical Association*, **77**, 416–424.
- Mosteller, F. and Tukey, J.W. (1977). *Data Analysis and Regression*, Addison-Wesley, Reading.
- Nadarajah, S. (2005). A generalized normal distribution, *Journal of Applied Statistics*, **32**(7), 685–694.

Olmos, N.M., Varela, H., Gomez, H.W. and Bolfarine, H. (2012). An extension of the half-normal distribution, *Statistical Papers*, **53**, 875–886.

Rogers, W.H. and Tukey, J.W. (1972). Understanding some long-tailed symmetrical distributions, *Statistica Neerlandica*, **26**, 211–226.

Wang, J. and Genton, M.G. (2006). The multivariate skew-slash distribution, *Journal of Statistical Planning and Inference*, **136**, 209–220.



A note on the mean remaining time to failure of the censored data under type-I and type-II censoring schemes

M.H. Poursaeed ¹

¹Department of Mathematics, Lorestan University, 465 Khoramabad, IRAN

Abstract: The most common case of censoring is what is referred to as right censored data. There are two well known right censoring schemes which are conventional type-I and type-II censoring. In this paper, we propose two expected values for measuring the remaining lifetime of censored data, under these censoring schemes, by which one can predict the the failure time of censored units. Several properties of the proposed measures are derived. We also propose a likelihood function, call it predicted likelihood, to estimate the unknown parameters of a statistical model under type-I and type-II censoring. It is shown that, in the case where the model is exponential, the estimator based on which maximizes predicted likelihood is sometimes more efficient than the maximum likelihood estimator based on classical method.

Keywords Mean residual life function, reliability function , hazard rate, likelihood function, exponential distribution.

Mathematics Subject Classification (2010): 99X99 99X99 99X99.

1 Introduction

There are several scenarios in life-testing and reliability experiments in which units that are subject to test are lost or removed from the experiment before failure. Such units are usually called the censored units. Two well known schemes of right censoring are conventional type-I and type-II censoring which are extensively studied in statistical and reliability literature. We refer the reader to Lawless (1982), Bain and Engelhardt(1991), Balakrishnan and Cohen (1991) and Gertsbakh (2000), among others. The cited authors, have considered lifetime studies in both parametric and non-parametric models in industrial and actuarial context. Assume that n units are placed on a life-test at time 0. Type-I censoring occurs when the individuals only observed for a pre-fixed time t , independent of the failure times. Based on this scheme, for each unit, we will only observe its failure if this occurs before t . The units with lifetime greater than t are called the censored units. Type-II censoring arise when only the m smallest lifetimes (from the n) are observed. Hence under this scheme, when n items are put on test the decision is made to terminate the test when the m th failure occurs.

¹Speaker's Name: M.H. Poursaeed poursaeed-mh@yahoo.com

Let T_1, \dots, T_n denote the lifetimes of the units. Assume that the T_i 's, $i = 1, \dots, n$ are independent and have a common distribution function F . Let also $T_{1:n}, \dots, T_{n:n}$ be the order statistics corresponding to T_i 's. In the present note we consider two conditional random variables ϕ and ψ as follows

$$\phi(t) = (T_{r:n} - t | T_{m:n} < t < T_{m+1:n}) \quad 0 \leq m < m+1 \leq r \leq n,$$

and

$$\psi(t) = (T_{r:n} - t | T_{m:n} = t) \quad 0 \leq m < m+1 \leq r \leq n,$$

with $T_{0:n} = 0$.

The random variables ϕ and ψ can be interpreted as follows.

- $\phi(t)$ is the residual lifetime of the censored data $T_{r:n}$, under the condition that, in type-I censoring scheme, before time t we have observed m failures and $n - m$ observations are censored, $r = m + 1, \dots, n$.
- $\psi(t)$ gives the residual lifetime of $T_{r:n}$, under the condition that, in type-II censoring scheme, the m th failure occurs at time t and the rest of observations are censored, $r = m + 1, \dots, n$.

The aim of this paper is to obtain the expected values of ϕ and ψ , i.e. $E(\phi(t))$ and $E(\psi(t))$, which we denote them by $m_n^{r,m}(t)$ and $M_n^{r,m}(t)$, respectively. Having these measures enable us to predict the failure time of the censored data under type-I and type-II censoring. In section 2, we concentrate on $m_n^{r,m}(t)$, as a adynamic function of t , and show that how it can be calculated based on the distribution function. It is shown that the functional form of $M_n^{r,m}(t)$, which is only a function of the failure time of the m th failure, is the same as that of $m_n^{r,m}(t)$. Several properties of $m_n^{r,m}(t)$ (and consequently of $M_n^{r,m}(t)$) are studied. It is proved, among others, that $m_n^{r,m}(t)$ is a monotone function of r , m and n , respectively. When the distribution function F has a monotone hazard rate, we get that $m_n^{r,m}(t)$ is also monotone function of time. We also show that $m_n^{r,m}(t)$ uniquely determines the underlying distribution function F . In Section 3 of the paper, we use $m_n^{r,m}(t)$ and $M_n^{r,m}(t)$, in an inferential point of view, to estimate the unknown parameters of a distribution. Using these measures we introduce two likelihood functions, call them *predicted likelihood functions*. It will be shown, as an example, that in the case where the parent distribution is exponential with mean θ , the use of predicted likelihoods give estimators which are sometimes more efficient than the classical maximum likelihood estimators.

2 The mean remaining time to failure of the censored data

In this section we consider $m_n^{r,m}(t)$ as a dynamic function of t and obtain its functional form. To this end, one can show that for $r = m + 2, \dots, n$, $m = 0, 1, \dots, r - 1$ and $0 < s$,

$$P(T_{r:n} - t > s, T_{m:n} < t < T_{m+1:n}) = \binom{(\cdot)}{n}, m (F(t))^m \sum_{i=n-r+1}^{n-m} \binom{(\cdot)}{n} - m, i (\bar{F}(t+s))^i (F(t+s) - F(t))^{n-m-i}.$$

On the other hand we can easily verify that

$$P(T_{m:n} < t < T_{m+1:n}) = \binom{(\cdot)}{n}, m (F(t))^m (\bar{F}(t))^{n-m}.$$

Hence, using the results above, we obtain for $r = m + 2, \dots, n$ and $0 < s$.

$$\begin{aligned} P(T_{r:n} - t > s | T_{m:n} < t < T_{m+1:n}) &= \frac{P(T_{r:n} - t > s, T_{m:n} < t < T_{m+1:n})}{P(T_{m:n} < t < T_{m+1:n})} \\ &= \sum_{i=n-r+1}^{n-m} \binom{(\cdot)}{n} - m, i \left(\frac{\bar{F}(t+s)}{\bar{F}(t)} \right)^i \left(1 - \frac{\bar{F}(t+s)}{\bar{F}(t)} \right)^{n-m-i}, \end{aligned}$$

where $\bar{F} = 1 - F$ denote the survival function of the units. One can also see that, using a different argument, that this last expression is valid for $r = m + 1$. Thus, the mean remaining time to failure function $m_n^{r,m}(t)$ can be obtained as

$$\begin{aligned} m_n^{r,m}(t) &= E(T_{r:n} - t | T_{m:n} < t < T_{m+1:n}) \\ &= \int_0^\infty P(T_{r:n} - t > s | T_{m:n} < t < T_{m+1:n}) ds \\ &= \int_0^\infty \sum_{i=n-r+1}^{n-m} n - m, i \left(\frac{\bar{F}(t+s)}{\bar{F}(t)} \right)^i \left(1 - \frac{\bar{F}(t+s)}{\bar{F}(t)} \right)^{n-m-i} ds \\ &= \int_t^\infty \sum_{i=n-r+1}^{n-m} n - mi \left(\frac{\bar{F}(s)}{\bar{F}(t)} \right)^i \left(1 - \frac{\bar{F}(s)}{\bar{F}(t)} \right)^{n-m-i} ds \quad r = m + 1, \dots, n. \end{aligned}$$

Remark 2.1. It can be shown, using the same method, the mean remaining time to failure $M_n^{r,m}(t)$, under the type-II censoring, has the same functional form of $m_n^{r,m}(t)$. That is, one can show that

$$M_n^{r,m}(t) = \int_t^\infty \sum_{i=n-r+1}^{n-m} n - mi \left(\frac{\bar{F}(s)}{\bar{F}(t)} \right)^i \left(1 - \frac{\bar{F}(s)}{\bar{F}(t)} \right)^{n-m-i} ds \quad r = m + 1, \dots, n. \quad (2.1)$$

It should be pointed out, however, that (although we assumed that $m_n^{r,m}(t)$ is a function of t) in censoring type-I t is fixed and m is a random variable whereas $M_n^{r,m}(t)$ is a function of t , the failure time of the m th failure, and m is fixed.

Although the form of $m_n^{r,m}(t)$ and $M_n^{r,m}(t)$, are similar to the mean residual lifetime function of a system (for more information, see Guess and Proschan (1988) and Poursaeed (2010)), they have different concepts.

In the sequel, we concentrate on $m_n^{r,m}(t)$ and obtain some of its properties. The results are also valid for $M_n^{r,m}(t)$.

Remark 2.2. The mean remaining time to failure $m_n^{r,m}(t)$ can also be represented as

$$m_n^{r,m}(t) = \int_t^\infty P(Y_t^u \geq n - r + 1) du$$

where Y_t^u is distributed as binomial with parameters $(n - m, \theta_t(u))$, where for $u > t$ $\theta_t(u) = \frac{\bar{F}(u)}{\bar{F}(t)}$.

The following theorem shows the behavior of $m_n^{r,m}(t)$ in terms of n , m and r .

Theorem 2.3. (a) For fixed m and r , $m_n^{r,m}(t)$ is a decreasing function of n .

(b) For fixed n and r , $m_n^{r,m}(t)$ is a decreasing function of m .

(c) For fixed n and m , $m_n^{r,m}(t)$ is an increasing function of r .

Proof:

(a) Some algebraic manipulations show that

$$P(Y_t^u \geq n - r + 1) - P(Z_t^u \geq (n + 1) - r + 1) = n - m - r + 1(\theta_t(u))^{n-r+1}(1 - \theta_t(u))^{r-m}$$

where Y_t^u and Z_t^u are distributed as binomial with parameters $(n - m, \theta_t(u))$ and $((n + 1) - m, \theta_t(u))$, respectively. From this, we conclude that $m_n^{r,m}(t)$ is a decreasing function of n . In particular when the distribution function F is strictly increasing, $m_n^{r,m}(t)$ is strictly decreasing function of n .

(b) Similar to Part (a) one can show that

$$P(Y_t^u \geq n - r + 1) - P(Z_t^u \geq n - r + 1) = n - m - 1 - r(\theta_t(u))^{n-r+1}(1 - \theta_t(u))^{r-m-1}$$

where Y_t^u and Z_t^u are distributed as binomial with parameters $(n - m, \theta_t(u))$ and $(n - (m + 1), \theta_t(u))$, respectively. This completes the proof of part (b).

(c) To prove the assertion of this part of theorem note that

$$\begin{aligned} m_n^{(r+1),m}(t) - m_n^{r,m}(t) &= E(T_{r+1:n} - T_{r:n} | T_{m:n} < t < T_{m+1:n}) \\ &\geq 0 \end{aligned}$$

Remark 2.4. Although, it is shown in Part (a) of Theorem 2.3 that for fixed r and m , $m_n^{r,m}(t)$ is decreasing in n , it can be shown that $m_n^{n,m}(t)$ is an increasing function of n . To see this note that

$$\begin{aligned} m_n^{n,m}(t) &= \int_t^\infty [1 - (1 - \frac{\bar{F}(s)}{\bar{F}(t)})^{n-m}] ds \\ &\leq \int_t^\infty [1 - (1 - \frac{\bar{F}(s)}{\bar{F}(t)})^{n+1-m}] ds \\ &= m_{n+1}^{n+1,m}(t). \end{aligned}$$

Remark 2.5. It is not hard to verify that the following result is true.

$$\begin{aligned} m_n^{r-1,m}(t) &= m_{n+1}^{r,m+1}(t) \\ &\leq m_{n+1}^{r,m}(t) \\ &\leq m_n^{r,m}(t) \\ &\leq m_{n+1}^{r+1,m}(t) = m_n^{r,m-1}(t) \end{aligned}$$

The hazard rate is a crucial concept in life studies. If the distribution function F has a density function f the hazard rate corresponding to F , which we denote here by $r(t)$ is defined as

$$r(t) = \frac{f(t)}{\bar{F}(t)}.$$

The following theorem proves that when the distribution function F has a monotone hazard rate then $m_n^{r,m}(t)$ is monotone.

Theorem 2.6. If the hazard rate function $r(t)$ is an increasing (decreasing) function of t , then $m_n^{r,m}(t)$, as a function of t , is a decreasing (increasing) function of t .

Proof: The assumption that $\frac{f(t)}{\bar{F}(t)}$ is a increasing (decreasing) function of t is equivalent to say that $\frac{\bar{F}(t+s)}{\bar{F}(t)}$ is an decreasing (increasing) function of t . Hence, for $t_1 < t_2$ we have

$$\begin{aligned} m_n^{r,m}(t_1) &= \int_0^\infty P(Y_{t_1}^{t_1+s} \geq n-r+1) ds \\ &= \int_0^\infty \int_0^{\frac{\bar{F}(t_1+s)}{\bar{F}(t_1)}} \frac{(n-m)!}{(n-r)!(r-m-1)!} x^{n-r} (1-x)^{r-m-1} dx \end{aligned}$$

$$\begin{aligned}
 &\geq (\leq) \int_0^\infty \int_0^{\frac{\bar{F}(t_2+s)}{\bar{F}(t_2)}} \frac{(n-m)!}{(n-r)!(r-m-1)!} x^{n-r} (1-x)^{r-m-1} dx \\
 &= \int_0^\infty P(Y_{t_2}^{t_2+s} \geq n-r+1) ds \\
 &= m_n^{r,m}(t_2).
 \end{aligned}$$

This completes the theorem.

The following theorem is an interesting result showing that the distribution function F can be recovered from $m_n^{r,m}(t)$.

Theorem 2.7. *Let F be a strictly increasing distribution function. Then, F can be uniquely characterized by $m_n^{r,m}(t)$ as follows.*

$$F(t) = 1 - \exp\left\{ -\frac{1}{(n-m)} \int_0^t \frac{1 + \frac{d}{dx} m_n^{n,m}(x)}{m_n^{r,m}(x) - m_n^{r,m+1}(x)} dx \right\} \tag{2.2}$$

Proof: The proof follows using the same steps as used by Asadi and Bayramoglu (2006) to prove their Theorem 1. It should be mentioned that the assumption that F is strictly increasing gives a sufficient condition to have $m_n^{r,m}(t) > m_n^{r,m+1}(t)$. This makes the integrant in (2.2) to be well defined.

3 Inference based on predicted likelihood

In this section, we assume that the statistical model F depends on some unknown parameters. For the sake of simplicity, we assume that F depends on one unknown parameter θ . To estimate the unknown parameter θ , the maximum likelihood method is the most common method in the statistical literature. The main advantage of the maximum likelihood method is that it can be adjusted for incomplete data such as censored data. As it was already mentioned, in type-II censoring the information available consists of the failure times of first m order statistics $T_{1:n} = t_1, \dots, T_{m:n} = t_m$ and the only information about the remaining $n - m$ observations is that they exceed t_m . Hence, under the assumption that the distribution function F has a density function f , the likelihood function, which we denote by $L(\theta)$, has the following form :

$$L(\theta) = \prod_{i=1}^m f(t_i; \theta) (\bar{F}(t_m; \theta))^{n-m} \tag{3.1}$$

In type-I censoring scheme, provided that there is at least one failure time before t , the likelihood function has the form:

$$L(\theta) = \prod_{i=1}^m f(t_i; \theta) (\bar{F}(t; \theta))^{n-m} . \tag{3.2}$$

In view of the results of Section 2, $m_n^{r,m}(t)$ and $M_n^{r,m}(t)$ predict the failure times of the remaining lifetime of the censored data in type-I censoring and type-II censoring, respectively. In other words, the quantities $m_n^{r,m}(t)+t$ and $M_n^{r,m}(t)+t$ give predictions of the failure times of the censored data under the mentioned censoring schemes. Using these predictions, we propose the following *predicted likelihoods* to estimate the parameter θ . Under type-I censoring, the *predicted likelihood* (PL) is proposed as

$$PL(\theta) = \prod_{i=0}^m f(t_i; \theta) \prod_{i=m+1}^n f(m_n^{i,m}(t) + t; \theta) \quad (3.3)$$

with $f(t_0; \theta) = 1$. In type-II censoring the PL is proposed as

$$PL(\theta) = \prod_{i=1}^m f(t_i; \theta) \prod_{i=m+1}^n f(M_n^{i,m}(t) + t; \theta). \quad (3.4)$$

It worth to mention that, in the above predicted likelihoods, the first product on the right hand sides evaluate the density at exact failure times of the first m failures and the second product evaluate the density at the predicted failure times of the remaining $n - m$ data. A maximum corresponding to *predicted likelihood* is called *maximum predicted likelihood estimate* (MPLE).

In the following we give some examples to show that how one can use PL method to estimate the unknown parameter θ of the exponential model. In particular, we show that under type-II censoring, for some values of m and n , the MPLE of θ is more efficient than the MLE based on the classical method when we compare the estimators in terms of their mean square error.

Example 3.1. Let F is exponential distribution with mean θ . Assume that under type-II censoring m failure times $m \geq 1$ are observed. Then it is well known that the MLE of θ is given by

$$\hat{\theta}_1 = \frac{(n-m)T_{m:n} + \sum_{i=1}^m T_{i:n}}{m}.$$

See, for example Gertsbakh (2000).

To obtain MPLE of θ first we compute $M_n^{r,m}(t)$. We have

$$M_n^{r,m}(t) = \theta \sum_{i=n-r+1}^{n-m} \sum_{j=0}^{n-m-i} n - \min - m - ij \frac{(-1)^j}{i+j}$$

Hence, The predicted likelihood is

$$\begin{aligned} PL(\theta) &= \left(\frac{1}{\theta}\right)^n e^{-\frac{\sum_{i=1}^m t_i + (n-m)t}{\theta}} \left(e^{-\sum_{r=m+1}^n \sum_{i=n-r+1}^{n-m} \sum_{j=0}^{n-m-i} n - \min - m - ij \frac{(-1)^j}{i+j}} \right) \\ &\propto \left(\frac{1}{\theta}\right)^n e^{-\frac{\sum_{i=1}^m t_i + (n-m)t}{\theta}} \end{aligned}$$

From this we get that the MPLE of θ is

$$\begin{aligned}\hat{\theta}_2 &= \frac{(n-m)T_{m:n} + \sum_{i=1}^m T_{i:n}}{n} \\ &= \left(1 - \frac{m}{n}\right)T_{m:n} + \frac{m}{n}\bar{T}\end{aligned}$$

where $\bar{T} = \frac{\sum_{i=1}^m T_{i:n}}{m}$. This shows that MPLE of θ is a weighted average of the mean of failure times of the m failed observations and the time of m th failure. It is also clear that $\hat{\theta}_2 < \hat{\theta}_1$ which in turn implies that $\hat{\theta}_2$ is not unbiased for θ . It is well known that (see, for example, Gertsbakh (2000)) $(n-m)T_{m:n} + \sum_{i=1}^m T_{i:n}$ is distributed as Gamma (m, θ) . Thus, one can conclude that the mean square errors of $\hat{\theta}_1$ and $\hat{\theta}_2$ are, respectively, given by

$$MSE(\hat{\theta}_1) = \frac{\theta^2}{m}$$

and

$$MSE(\hat{\theta}_2) = \theta^2 \left(\frac{m^2}{n^2} + \frac{m}{n^2} - \frac{2m}{n} + 1 \right).$$

This implies that for $m = 1$ we always have $MSE(\hat{\theta}_2) < MSE(\hat{\theta}_1)$ and for $m > 1$ it can be easily verified that for $n < \frac{m(m+1)}{m-1}$, $MSE(\hat{\theta}_2) < MSE(\hat{\theta}_1)$.

Table 1 shows the efficiency of $\hat{\theta}_2$ relative to $\hat{\theta}_1$ for some values of m and n which satisfy the condition $n < \frac{m(m+1)}{m-1}$. As the table indicates, for these choices of m and n , $\hat{\theta}_2$ is more efficient than of $\hat{\theta}_1$ and that the relative efficiency tends to 1 as n gets larger and m tends to n .

Example 3.2. If we assume that the observations are obtained based on type-I censoring, with prefixed t , from an exponential distribution with mean θ , then the MLE of θ is given by (see, for example, Bain and Engelhardt(1991)).

$$\hat{\theta}_1 = \frac{(n-m)t + \sum_{i=1}^m T_{i:n}}{m}.$$

If we use the predicted likelihood, it can be shown, similar to type-II censoring, that

$$\hat{\theta}_2 = \frac{(n-m)t + \sum_{i=1}^m T_{i:n}}{n}$$

But, it needs to be pointed out that m is random here. Let us, assume that $n = 2$, and that we observe $m = 1$ failure before t . On the other hand, under these conditions $m_2^{2,1}(t) = \theta$. Therefore

$$\begin{aligned}\hat{\theta}_1 &= T_{1:2} + t \\ \hat{\theta}_2 &= \frac{T_{1:2} + t}{2}.\end{aligned}$$

Table 1.

m	n	$\frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}$
2	3, 4, 5, 6	0.67, 0.75, 0.87, 1
3	4, 5, 6	0.75, 0.84, 1
4	5, 6	0.8, 0.89
5	6, 7	0.83, 0.91
6	7, 8	0.84, 0.94
7	8, 9	0.87, 0.95
8	9, 10	0.89, 0.96
9	10, 11	0.90, 0.97
10	11, 12	0.90, 0.97
11	12, 13	0.91, 0.98
12	13, 14	0.91, 0.98
13	14, 15	0.92, 0.98
14	15, 16	0.93, 0.98
15	16, 17	0.94, 0.99
16	17, 18	0.94, 0.99
17	18, 19	0.94, 0.99
18	19, 20	0.95, 0.99
19	20, 21	0.95, 0.99
20	21, 22	0.95, 0.99

It is not hard to see that

$$MSE_1 = \frac{\theta^2}{2} - t\theta + t^2$$

and

$$MSE_2 = \frac{5\theta^2}{8} - \frac{3t\theta}{4} + \frac{t^2}{4}.$$

Comparing the MSE's, one can see in this case that for $t > \frac{\theta(1+\sqrt{7})}{6}$, $MSE_2 < MSE_1$. That is, $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$. for

Remark 3.3. It has to be pointed out that in type-I censoring to have an MLE for θ we must observe at least one failed item before time t , i.e. $m > 0$. Otherwise, The maximum likelihood method does not work. However, using the predicted likelihood method proposed here may enables one to have an MPLE for the parameters of the distributions. Note that, in this case, we have

$$PL(\theta) = \prod_{i=1}^n f(m_n^{i,0}(t) + t; \theta). \quad (3.5)$$

To illustrate more let us consider the exponential distribution an assume that n items are put on the test and for a prefixed time t we have not observed any failure before t . Thus we have

$$\begin{aligned}
m_n^{r,0}(t) &= \sum_{i=n-r+1}^n ni \int_t^\infty \left(\frac{\bar{F}(u)}{\bar{F}(t)} \right)^i \left(1 - \frac{\bar{F}(u)}{\bar{F}(t)} \right)^{n-i} du \\
&= \sum_{i=n-r+1}^n \sum_{j=0}^{n-i} nin - ij(-1)^j \int_t^\infty (e^{-\frac{(u-t)}{\theta}})^{i+j} du \\
&= \theta \sum_{i=n-r+1}^n \sum_{j=0}^{n-i} nin - ij \frac{(-1)^j}{i+j}
\end{aligned}$$

This yields

$$PL(\theta) = \frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum_{r=1}^n (t+\theta \sum_{i=n-r+1}^n \sum_{j=0}^{n-i} nin - ij \frac{(-1)^j}{i+j})}$$

Which in turn implies, after some manipulations, that the MPLE of θ is

$$\hat{\theta} = t.$$

Remark 3.4. When the observations are obtained from the uniform distribution on $(0, \theta)$ and also based on type-I censoring with prefixed t , it can be shown that, the maximum likelihood method still does not work for $m=0$. While the MPLE of θ is $\hat{\theta} = t$.

References

- Asadi, M and Bairamov, I (2006). On the mean residual life function of the k -out-of- n systems at the system level. *IEEE Transactions in Reliability*, 55: 314-318
- Bain, LJ and Engelhardt, M (1991). *Statistical Analysis of Reliability and Life-Testing Models*, Second edition, Marcel Dekker, New York.
- Balakrishnan, N and Cohen, AC (1991). *Order Statistics and Inference: Estimation Methods*, Academic Press, San Diego.
- Gertsbakh, I (2000). *Reliability Theory With Applications to Preventive Maintenance*, Springer.
- Guess F, Proschan F (1988) *Mean residual life: theory and applications*. In: Krishnaiha PR, Rao CR (eds) Handbook of statistics, vol 7. North-Holland, Amsterdam, pp 215-224
- Lawless, JF (1983). *Statistical Models and Methods for Lifetime Data*. New York; Wiley.
- Poursaeed MH (2010) A note on the mean past and the mean residual life of a $(n-k+1)$ -out-of- n system under multi monitoring. *Statistical Papers* , 51: 409-419



A Long Memory Linear Arch Model and R/S Statistic for Estimating its Long Memory Parameter

Rohollah Ramezani¹, Seyed Abolfazl Mousavi²

^{1,2}Department of Statistics, Faculty of Mathematics and Computer Science, University of Damghan, Damghan, Semnan, Iran

Abstract: Usually, it is assumed that the variance is invariant in the examination of time series data. But the variance is not invariant in applications. Therefore, various models are suggested so that the variance is expressed as a function of observation. one of the well-know models which often is used is an autoregressive conditional heteroscedasticity (*ARCH*) model that is used in economic fields such as foreign exchange rate, stock prices. In this article assumed that r_t is conditionally heteroscedasticity nonlinear (*ARCH*) process for which the conditional variance of the observable sequence r_t is the square of an inhomogeneous linear combination of r_s , $s < t$, we give conditions under which, for integers $l \geq 2$, r_t^l has long memory autocorrelation and normalized partial sums of r_t^l convergence to fractional Brownian motion. In the next we consider estimator for a long memory parameter in introduced model.

Keywords conditional heteroskedasticity, ARCH models, long memory, R/S estimator.

Mathematics Subject Classification (2010): 62M10 37M10.

1 Introduction

Engle (1982) and Bollerslev (1986) presented ARCH and GARCH sequences for the first time and have been used extensively to model financial time series, such as asset returns and exchange rates. A common property of *ARCH*(p) and *GARCH*(p, q) sequences is that they are defined by finite recursions and their autocorrelations decrease very rapidly, implying short memory behavior of these sequences. The short memory behavior holds even for *ARCH*(∞) models defined by the infinite recursion. The *ARCH*(∞) is defined following:

$$r_t = \varepsilon_t \sigma_t, \quad \sigma_t^2 = \alpha + \sum_{j=1}^{\infty} \beta_j r_{t-j}^2, \quad t \in Z$$

For constants $\alpha > 0, \beta_j \geq 0$ (to ensure that $\sigma_t^2 > 0$), where the β_j also satisfy some summability condition, easily achieved in both *ARCH*(p) model of Engle (1982) (where in $b_j = 0, j > p$) and

²Seyed Abolfazl Mousavi : mousavi.stat@gmail.com

its GARCH extension of [Bollerslev \(1986\)](#). However, these latter models imply exponential decay in autocorrelations of the r_t^2 , where empirical evidence has frequently suggested a much greater degree of persistence, possibly consistent with long memory in r_t^2 , where autocorrelation are not summable [see [Ding et al. \(1993\)](#)]. A model which describes the long memory behavior was suggested by [Giraitis et al. \(2009\)](#). This model is called $LMLARCH(\infty)$ (Long Memory Linear ARCH) and defined by the following equations:

$$r_t = \varepsilon_t \sigma_t, \quad \sigma_t = \alpha + \sum_{j=1}^{\infty} \beta_j r_{t-j}, \tag{1.1}$$

2 Properties of $LMLARCH(\infty)$ model

We consider first the following assumptions.

Assumption 1. (i) (1.1) hold.

(ii) $\{\varepsilon_t\}$ is a sequence of iid random variables with zero mean and unit variance.

(iii) $\alpha \neq 0$

Assumption 2. $\beta^2 = \sum_{j=1}^{\infty} \beta_j^2 < 1$.

Let F_t be the σ -field of events by $\varepsilon_s, s \leq t$.

Theorem 2.1. *Let Assumption 1 hold. Then a covariance stationary F_{t-1} measurable solution $\sigma_t, t \in \mathbb{Z}$ of (1.1) exists if and only if Assumption 2 holds, in which case, for $t \in \mathbb{Z}$, we have the Volterra expansion*

$$\sigma_t = \alpha + \alpha \sum_{k=1}^{\infty} \sum_{j_1, \dots, j_k=1}^{\infty} \beta_{j_1} \dots \beta_{j_k} \varepsilon_{t-j_1} \dots \varepsilon_{t-j_1-\dots-j_k}$$

and

$$E(\sigma_t) = \alpha, \quad Cov(\sigma_0, \sigma_t) = \frac{\alpha^2}{1 - \beta^2} \sum_{j=1}^{\infty} \beta_j \beta_{j+t}. \tag{2.1}$$

From (2.1), we can also write

$$Corr(\sigma_0, \sigma_t) = \frac{\sum_{j=1}^{\infty} \beta_j \beta_{j+t}}{\beta^2} \tag{2.2}$$

Which we recognize as the usual formula for the autocorrelation function in terms of wold decomposition weights. we can thus control the memory of σ_t by choice of β_j . We introduce the next assumption.

Assumption 3. assume that

$$\beta_j \sim c j^{d-1} \quad 0 < c < \infty, \quad 0 < d < 1/2, \tag{2.3}$$

where " \sim " indicates that the ratio of left and right sides tends to 1.

Corollary 2.2. *Let Assumptions 1-3 hold. Then*

$$\text{Cov}(\sigma_0, \sigma_t) \sim kt^{2d-1}, \quad k > 0. \quad (2.4)$$

The proof is standard from the long memory literature, using (2.3) and (2.1). An example of b_j satisfying assumption 3 is

$$b_j = c \frac{\Gamma(j+d)}{\Gamma(j+1)} \quad (2.5)$$

Which is proportional to the moving average weight in a standard fractional ARIMA(0,d,0) model. It will be found that, for integer $l \geq 2$, r_t^l has autocorrelations decaying at the same rate as those of σ_t when σ_t has long memory, and that the normalized partial sums of r_t^l (like those of σ_t) converge to fractional Brownian motion. Notice that typically σ_t is unobservable, where r_t is observable, so that its autocovariance can likely be consistently estimated under suitable conditions. To establish the properties of r_t^l , we impose also:

Assumption 4. ε_t has finite $2l^{\text{th}}$ moments such that

$$(4^l - 2l - 1)\mu_{2l}^{1/l}\beta^2 < 1, \quad \mu_j = E(\varepsilon_0^j). \quad (2.6)$$

For given l , (2.6) is a tighter restriction on β than Assumption 2, while (2.6) becomes more stringent as l increase, since $(4^l - 2l - 1)$ and $\mu_{2l}^{1/l}$ are increasing functions, so that (2.6) holds also for $j < l$. When ε_t is Gaussian $\mu_{2l} = (2l - 1)(2l - 3) \dots 3 \cdot 1$, thought in this case is likely that the factor $(4^l - 2l - 1)$ can be reduced, (2.6) being only a sufficient condition for the following results.

Theorem 2.3. *Let Assumptions 1,2,3 and 4 hold. Then, for $k=2, \dots, l$*

$$\text{Cov}(r_0^k, r_t^k) \sim \lambda_j^2 t^{2d-1} a s t \rightarrow \infty, \quad \lambda_j = m_j E(r_0^j) \quad (2.7)$$

where m is a positive constant.

The following Theorem represents the limit distribution of normalizes partial The following Theorem represent the limit distribution of normalizes partial sums of the r_t^l . Let $\{W_\gamma(t), t \geq 0\}$, the fractional Brownian motion with parameter d ; ($0 < d < 1/2$), is a Gaussian process with mean 0 and covariance

$$E[W_H(t)W_H(s)] = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |s-t|^{2H}), \quad (2.8)$$

See, for example, Samorodnitsky et al. (1994) [Chapter 7]. Let $[]$ denote integer part and \implies the convergence of finite-dimensional distributions.

Theorem 2.4. *Under Assumptions 1,2,3 and 4, for $j=2,\dots,l$, as $N \rightarrow \infty$,*

$$N^{-(2d+1)/2} \sum_{s=1}^{[Nt]} (r_s^j - E(r_s^j)) \Rightarrow k_d \lambda_j W_{d+1/2}(t), \quad t \geq 0. \tag{2.9}$$

where

$$k_d = \left\{ \frac{2}{(2d)(2d+1)} \right\}^{1/2} \tag{2.10}$$

3 The Estimator

Rescaled range or R/S analysis which introduced by [Hurst \(1951\)](#) and subsequently refined by [Mandelbrot et al. \(1979\)](#) has been used in this paper for estimating long memory parameter d in [\(1.1\)](#) and [\(2.3\)](#). The R/S statistic is defined as $\hat{R}_N/\hat{\sigma}_N$ where

$$\hat{R}_N = \max_{1 \leq k \leq N} \sum_{j=1}^k (Y_j - \bar{Y}_N) - \min_{1 \leq k \leq N} \sum_{j=1}^k (Y_j - \bar{Y}_N) \tag{3.1}$$

is the range and

$$\hat{\sigma}_N^2 = \frac{1}{N} \sum_{j=1}^N (Y_j - \bar{Y}_N)^2 \tag{3.2}$$

is standard variance estimator and \bar{Y}_N is the sample mean.

by [Theorem 2.4](#) imply that

$$\frac{\hat{R}_N}{N^{1/2+d}} \rightarrow^d c_d \left\{ \max_{0 \leq t \leq 1} W_{1/2+d}^0(t) - \min_{0 \leq t \leq 1} W_{1/2+d}^0(t) \right\}, \tag{3.3}$$

where

$$W_{1/2+d}^0(t) = W_{1/2+d}(t) - tW_{1/2+d}(1)$$

is a fractional Brownian bridge, cf. [2.8](#). It is equally easy to verify that

$$\hat{\sigma}_N^2 \rightarrow^P \text{Var}(Y_1) \tag{3.4}$$

Indeed,

$$\hat{\sigma}_N^2 = \frac{1}{N} \sum_{j=1}^N (Y_j^2 - EY_j^2) + (EY_1^2 - \bar{Y}_N^2). \tag{3.5}$$

By the Volterra representation Y_j^2 can be written as $Y_j^2 = f(\varepsilon_j, \varepsilon_{j-1}, \dots)$ where f is a measurable function. Since $\{\varepsilon_j\}$ is an ergodic sequence. This implies ergodicity of Y_j^2 . Under assumptions of

Theorem 2.4 $EY_2^j < \infty$. Therefore the first term in 3.5 tends to zero. By the same argument as above $\{Y_j\}$ is ergodic as well, and therefore $\bar{Y}_N \Rightarrow EY_1$. Hence the second term in 3.5 tends to $VarY_1$.

Combining 3.3 and 3.4, we see that as $N \rightarrow \infty$

$$\frac{1}{N^{1/2+d}} \frac{\hat{R}_N}{\hat{\sigma}_N} \rightarrow^d \frac{c_d \{ \max_{0 \leq t \leq 1} W_{1/2+d}^0(t) - \min_{0 \leq t \leq 1} W_{1/2+d}^0(t) \}}{(VarY_1)^{1/2}} = R_d. \quad (3.6)$$

Relation 3.6 forms a theoretical foundation for the R/S method. Taking logarithms of both sides, we obtain a heuristic identity

$$\log\left(\frac{\hat{R}_N}{\hat{\sigma}_N}\right) \approx \left(\frac{1}{2} + d\right) \log N + \log R_d, \quad as n \rightarrow \infty,$$

which can also be written as

$$\hat{d}_{R/S} - d = O_P\left(\frac{1}{\log N}\right) \quad with \quad \hat{d}_{R/S} = \frac{\log(\hat{R}_N/\hat{\sigma}_N)}{\log N} - \frac{1}{2},$$

and which shows that $1/2 + d$ can be interpreted as the slope of a regression line of $\log(\frac{\hat{R}_N}{\hat{\sigma}_N})$ on $\log N$ with random intercept $\log R_d$. The point of the R/S analysis is to consider many subsamples of varying size N from a given sample Y_1, \dots, Y_N in order to obtain many points which are used to estimate the slope of the regression line, see for example Mandelbrot et al. (1979). The technical details of the implementation of this procedure are described in next Section.

4 Simulations

We consider the following data generating process (DGP):

$$r_k = \varepsilon_k \sigma_k, \quad \sigma_k = \alpha + \sum_{j=1}^{\infty} \beta_j r_{k-j}, \quad k = 1, \dots, N, \quad \varepsilon_k \sim N(0, 1). \quad (4.1)$$

Two sample sizes are considered, $N = 3000, N = 6000$. To reduce initialization effects, which are particularly strong for long range dependent sequences, we generate for each simulation a series $r_{-\tau}, r_{-\tau+1}, \dots, r_{-1}, r_0, r_1, \dots, r_N, \tau = 12000$, where the pre-sample observations $r_{-\tau}, r_{-\tau+1}, \dots, r_{-1}, r_0$ are recursively used for initializing the process, and are discarded afterwards. We truncate the infinite order lag polynomial $\beta(L)$ at the order 5000 to take into account the dependence between very distant observations. The sequence of innovations ε_k from standard Gaussian random variable is generated by using the Box-Muller method.

We generate 5000 independent samples. Each sample of N observation is subdivided into B adjacent

and non-overlapping blocks of observations of equal size $[N/B]$. We then obtain a grid $t_1 = 1, t_2 = [N/B] + 1, \dots, t_i = (i - 1)[N/B] + 1, \dots, t_B = N - [N/B] + 1$. For each point of the sequence $\{t_i\}_{i=1}^B$ we define a sequence of K increasing nested blocks with origin t_i , i.e., $\{[t_i, t_i + k_j]\}_{j=1}^K$ such that $t_i + k_j \leq N$, the sequence of K steps $\{k_j\}_{j=1}^K$ is given by a logarithmic grid. Interested readers are referred to [Beran \(1994\)](#), p. 84 – 85, for more details on the pox-plot based estimators. In our simulations, the minimum value of K is set to 40 and the number of blocks B is set to 40.

We calculate the R/S statistics for each interval $\{[t_i, t_i + k_j]\}_{i=1}^B\}_{j=1}^K$ and obtain the sequences $\{\{R/S(t_i, k_j)\}_{i=1}^B\}_{j=1}^K$. We plot the logarithm of the statistics $\log(R/S(t_i, k_j))$ against $\log(k_j)$ and then obtain a 'pox-plot'. Let \hat{b} be the slope of the least-squares regression line fitted to these pox-plots. Then $\hat{d}_{R/S} = \hat{b} - 1/2$.

Table 1: Estimation results for the $LMLARCH(\infty)$ process by using of R/S statistic

d	Model A		Model B	
	3000 obs	6000 obs	3000 obs	6000 obs
0.05	0.0495	0.0479	0.0434	0.0414
0.10	0.1102	0.1093	0.0597	0.0615
0.15	0.1470	0.1527	0.1088	0.1158
0.20	0.1887	0.1959	0.1697	0.1796
0.25			0.2251	0.2357
0.30			0.2661	0.2750
0.35			0.2880	0.2943

We have chosen two parameterizations of the moving average form of a $FARIMA$ process. For two models, the parameter α is set to 0.10.

(I) Model A. The coefficients β_j are those of the moving average representation of a $FARIMA(0, d, 0)$ process, that is, the coefficients of the fractional difference operator $(1 - L)^{-d}$. Thus, $\beta_j = b_j$, with the b_j defined recursively by $b_0 = 1, b_1 = d, b_j = b_{j-1}(j - 1 + d)/j, j > 1$.

(II) Model B. The coefficients of this DGP are those of the MA representation of a $FARIMA(1, d, 0)$ process with AR polynomial equal to $1 - \phi L$, that is, of the filter $(1 - \phi L)^{-1}(1 - L)^{-d}$. Thus, $\beta_1 = b_1 + \phi, \beta_j = \sum_{k=0}^{j-1} \phi^k b_{j-k}, j > 1$.

For the Model A, condition 2.6 is satisfied if $d < 0.1865$. If this condition is not satisfied, there is a large systematic bias for the pox-plot based estimators. For that reason, we do not report the estimates for $d > 0.20$.

For Models B, we choose $\theta = 0.20$ and $\phi = -0.20$. These values ensure that condition 2.6 is satisfied for $d > 0.1865$. The coefficients β_j depend on d , but also on the parameters θ and ϕ . If the first

elements of the sequence of the β_j are small, there is a systematic negative bias, that is, the parameter d is underestimated. This negative bias is quite large for small values of d , and becomes smaller when $d \in (0.20, 0.375)$, and increases for $d > 0.375$. It is well-known, see Mandelbrot et al. (1979), that for linear long-memory time series the R/S estimator overestimates d for small d and underestimates it for d close to 0.5.

Conclusion

In this paper, a long memory linear ARCH (LM LARCH) Model has been introduced. Base on assumptions 1 to 4, it is shown that $l \geq 2, r_t^l$ has the long memory property, where l is integer and normalized partial sums of r_t^l converge to fractional Brownian motion. In order to estimate the long memory parameter, the R/S estimator has been introduced.

References

- Beran, J. (1994). *Statistics for Long-Memory Processes*, Chapman and Hall, New York.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **31**, 307–327.
- Ding, Z., Granger, C.W.J. and Engle, R. F. (1993). A long memory property of stock market returns and a new model, *J. Empirical Finance*, **1**, 83–106.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1008.
- Giraitis, L., Remigijus, L. and Surgailis, D. (2009). *Handbook of Financial Time Series*, Springer.
- Hurst, H. (1951). Long term storage capacity of reservoirs, *Transac. Amer. Soc. Civil Engrs*, **116**, 770–799.
- Mandelbrot, B. B. and Taqqu, M. S. (1979). Robust R/S analysis of long run serial correlation, *42nd Session of the International Statistical Institute*, **2**, 69–90.
- Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes*, Chapman and Hall, New York.



Periodically Correlated Multi-component Locally Stationary Processes

S. Rezakhah¹, N. Modarresi²

¹Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Iran.

²Department of Mathematics and computer science, Allameh Tabataba'i University, Tehran, Iran.

Abstract: We introduce a new class of non-stationary processes, say periodically correlated locally stationary (PC-LS) process. Assume that $0 = s_0 < s_1 < \dots$ and $B_j = (s_{j-1}, s_j]$, for $j \in \mathbb{N}$. Let $X^{ls}(t) = \sum_{j=1}^{\infty} X_j^{ls}(t)I_{B_j}(t)$ where $X_j^{ls}(t)$ is a mixture of two stationary processes with exponentially convex weights. By this, we provide $X^{ls}(t)$ as a multi-component locally stationary process. Also we consider $\{X_j^p\}$ as a sequence of periodically correlated random variables. We define an orthogonally scattered random measure M_j on subsets of B_j by $X_j^p = M_j(B_j)$ and set $X_j^p(t) = M_j(s_{j-1}, t]$ for $t \in B_j$ with some special correlation. Then $X^p(t) = \sum_{j=1}^{\infty} X_j^p(t)I_{B_j}(t)$ is a continuous time periodically correlated process which we study its spectral representation. Finally we assume that $X^{ls}(t)$ and $X^p(t)$ are independent and define $X(t) = X^{ls}(t) + X^p(t)$ as a certain multi-component PC-LS process which has both periodically correlated and locally stationary properties. The covariance structure and the time dependent spectral representation of such a process are characterized.

Keywords Periodically correlated; Spectral representation; Multi-component locally stationary processes; exponentially convex covariance.

1 Introduction

Spectral analysis of stochastic processes such as stationary and periodically correlated (PC) processes have a long history with interest in both theory and applications. These processes are interesting for engineers due to their applications in signal processing and communication systems, specially in network traffic. The spectral theory of these processes is primarily based on spectral representations. Properties of stationary processes are well known and have been used extensively in analyzing system performance. Recall that a stationary process is one where all finite dimensional distributions are invariant to shifts in time. Non-stationary processes have extensive applications where random

¹S. Rezakhah : rezakhah@aut.ac.ir

fluctuations change in time or space. If a process is slowly varying and if the interval is short enough, then the process can be approximated in some sense by a stationary one. Recently a large amount of work has been devoted to time series analysis, with the focus placed on locally stationary (LS) and periodically correlated processes which give the plausible description of real world. LS processes in the sense of Silverman [Silverman \(1957\)](#) can be used to model systems where they behave as a function of time. He presented a relation between the covariance and the spectral density of LS processes which constitutes a natural generalization of the Wiener-Khintchine relations. Berman [Berman \(1974\)](#) introduced a class of LS Gaussian processes with index α which allows minor fluctuations of dependence at the global scale and at the same time keeps the stationary structure at the local time. Dehay [Dehay and Loughani \(1994\)](#) presented the notion of locally harmonizable process, a large class of non-stationary processes containing LS and harmonizable. A new class of LS processes where their spectral structure varies smoothly over time is introduced by Dahlhaus. Mallat et al. [Wahlberg and Hansson \(2007\)](#) showed that the covariance operator of a LS process has approximate eigenvectors that are local cosine functions. They modeled these processes with pseudo differential operators that are time-varying convolutions. Oxley et. al [Oxley, et. al. \(2000\)](#) proposed another definitions of LS processes and presented their properties and relationships to the stationary ones. They also showed that the signals that arise in Air Force applications typically has noise that can be modeled as a LS process. Exponentially convex stochastic processes and exponentially convex covariances have been studied by Loeve [Loeve \(1965\)](#), among others. He found the spectral representation of the exponentially convex process. Also Michalek obtained a spectral representation for these processes by enlarged the index to C (complex numbers). He also derived an inversion formula for the spectral measure. Another class of non-stationary processes are PC processes which have periodic structure. Theses processes are a class of processes which are in general non-stationary but exhibit many of the properties of stationary processes. They have been used as models of meteorology, radio physics and communications engineering. This class was introduced by Gladyshev [Gladyshev \(1961\)](#), who studied the structure of the covariance function and provided an interesting spectral representation. Hurd, Miamee, Makagon cited in Gardner [Gardner \(1994\)](#), demonstrated various applications of PC processes in science and engineering. In this paper we are concerned with a certain non-stationary process called, multi-component periodically correlated locally stationary (PC-LS) process. Let $0 = s_0 < s_1 < \dots$ and $B_j = (s_{j-1}, s_j]$, $j \in \mathbb{N}$. By this, we partition the positive index space into an infinite number of disjoint intervals B_j . Let $X(t) = X^{ls}(t) + X^p(t)$ represent a stochastic process, where $X^{ls}(t) = \sum_{j=1}^{\infty} X_j^{ls}(t)I_{B_j}(t)$ and $X_j^{ls}(t)$ is a multi-component LS process which is an exponentially convex mixture of two stationary processes and its covariance is shown in subsection 3.2 and 3.3. $X^p(\cdot)$ is a discrete time PC process that $X^p(t) = \sum_{j=1}^{\infty} X_j^p(t)I_{B_j}(t)$ in which $\{X_j^p(t)\}$ is a sequence of periodically correlated processes and $X_j^p(t) = M_j(s_{j-1}, t]$ where M_j for $j \in \mathbb{N}$ is an orthogonally scattered random measure on Borel field

of subsets of B_j . Then $X(\cdot)$ is a certain multi-component PC-LS process. More precisely this paper is organized as follows. In section 2, we study the general framework and preliminaries of discrete time periodically correlated, exponentially convex function and locally stationary processes. The harmonizability property of such processes like stationary and PC processes are exhibited in this section too. Spectral representation of continuous time PC processes have not been characterized in literature. We find the covariance function and the spectral representation of such continuous time multi-component PC-LS process as a new class of non-stationary process in section 3.

2 Preliminaries

We review the spectral theory of unitary operators in subsection 2.1. In subsection 2.2 we study the properties of stationary and periodically correlated processes and their spectral representations. We present some definitions of locally stationary and exponentially convex processes in subsection 2.3 and 2.4 respectively.

2.1 Spectral theory of unitary operators

One of the classical results of operator theory is the spectral theorem. We introduce the notion of spectral measure and briefly discuss their properties. We proceed to define the spectral integral as an operator. At last we represent the unitary operator and the time varying spectral representation of a PC process.

Throughout this subsection we work with a measurable space (Ω, \mathcal{F}) consisting of a set Ω and a σ -algebra \mathcal{F} of its subsets. For more details about the following results one could refer to Hurd and Miamee [Hurd and Miamee \(2007\)](#).

Definition 2.1. *A function Q defined on σ -algebra \mathcal{F} of subsets of Ω whose values are orthogonal projections in a Hilbert space \mathcal{H} is called a spectral measure, if $Q(\Omega) = I$, the identity operator, and for any sequence (M_n) of disjoint sets in \mathcal{F} we have $Q(\cup_{n=1}^{\infty} M_n) = \sum_{n=1}^{\infty} Q(M_n)$.*

The following proposition lists several useful properties of spectral measure from [Hurd and Miamee \(2007\)](#) without presenting the proofs.

Proposition 2.2. *If Q is an orthogonally scattered spectral measure on \mathcal{F} , where $Q(\emptyset) = 0$ we have the following properties*

(a) *Modular: for any two sets M, N in \mathcal{F}*

$$Q(M \cup N) = Q(M) + Q(N) - Q(M \cap N);$$

(b) *Multiplicative: for any two sets M, N in \mathcal{F}*

$$Q(M \cap N) = Q(M)Q(N).$$

Definition 2.3. *A unitary operator on a Hilbert space \mathcal{H} is a linear operator U from \mathcal{H} onto \mathcal{H} for which $\langle Ux, Uy \rangle = \langle x, y \rangle$ for every $x, y \in \mathcal{H}$, that is, unitary operators are linear and preserve inner products.*

Definition 2.4. *The spectral integral $\int f(\lambda)Q(d\lambda)$ as an operator $A(f)$ such that $A : \mathcal{H} \rightarrow \mathcal{H}$ and for any pair $x, y \in \mathcal{H}$*

$$\int f(\lambda)d\langle Q(\lambda)x, y \rangle = \langle A(f)x, y \rangle.$$

Theorem 2.5. *For any unitary operator U on a Hilbert space \mathcal{H} there exists a unique spectral measure Q on the Borel subsets of $[0, 2\pi)$ such that $U = \int_0^{2\pi} e^{i\lambda}Q(d\lambda)$, and for any integer t*

$$U^t = \int_0^{2\pi} e^{i\lambda t}Q(d\lambda).$$

Proposition 2.6. *A second order stochastic sequence X_t is PC with period T if and only if for every $t \in \mathbb{Z}$, there exists a unitary operator V^T and a periodic sequence (process) P_t taking values in $\mathcal{H}_X = \overline{\mathcal{L}_X}$ (closure of the \mathcal{L}_X) and $\mathcal{L}_X = sp\{X_t, t \in \mathbb{Z}\}$ for which*

$$X_t = V^t[P_t]$$

where $V = \int_0^{2\pi} e^{i\lambda/T}Q(d\lambda)$, $V^T = U$ and Q is the spectral measure defined in Definition 2.1.

2.2 Spectral representation of PC processes

We present definitions of stationary and PC processes in this subsection and exhibit the spectral representation of such processes and their covariance functions. For the following definitions and properties of PC and harmonizable processes, we refer to [Hurd and Miamee \(2007\)](#).

Definition 2.7. *A random process $X(t)$ taking values in the L_2 random variables of a probability space (Ω, \mathcal{F}, P) and indexed on \mathbb{Z} or \mathbb{R} is called periodically correlated if there exists some $T > 0$ in \mathbb{Z} or \mathbb{R} respectively, such that*

$$\mu(t) = E[X(t)] = \mu(t + T), \quad R_X(t, s) = R_X(t + T, s + T)$$

for every t, s in \mathbb{Z} or \mathbb{R} , $R_X(t, s) = E[(X(t) - \mu(t))\overline{(X(s) - \mu(s))}]$. The smallest such T will be called period of $X(t)$. For discrete time, it is required that $T > 1$, otherwise the process is stationary. For continuous time, also the process requires continuity of the correlation function.

The notion of a harmonizable process can be motivated from the fact that every wide sense stationary process has an integral spectral representation of the form

$$X(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega) \tag{2.1}$$

where the frequency indexed random process $Z(\omega)$ has orthogonal or uncorrelated increments and the equality is in L_2 of the probability space (Ω, \mathcal{F}, P) . The notion of harmonizable processes preserves the spectral representation (2.1) but the increments of $Z(\omega)$ need not be orthogonal.

We turn to spectral integral representation for PC sequences.

Theorem 2.8. *A second order stochastic sequence X_t is PC with period T if and only if there exists a time dependent spectral measure $\xi(\cdot, t) = \xi(\cdot, t + T)$ on the Borel subsets of $[0, 2\pi)$ that is orthogonally scattered in the sense that $\langle \xi(A, s), \xi(B, t) \rangle = 0$ for every $s, t \in \mathbb{Z}$ whenever $A \cap B = \emptyset$ and such that for all $t \in \mathbb{Z}$*

$$X_t = \int_0^{2\pi} e^{i\lambda t} \xi(d\lambda, t)$$

where the time dependent spectral measure $\xi(\cdot, t)$ is defined through the application of the spectral measure to the vector P_t as for any Borel set A , we have $\xi(A, t) = Q(A)P_t$, where P_t is introduced in Proposition 2.2.

2.3 Locally stationary processes

We present the concept of locally stationary process, which is a generalization of stationary random process and is introduced by Silverman [Silverman \(1957\)](#). The definition of exponentially convex process and its spectral representation is given too. Let $X(t)$ be a random process (in general complex), where the real parameter t lying in some index set I , which is a closed interval or the infinite real line \mathbb{R} . We assume that, for all $t \in I$, the second moment of $X(t)$ exists and the first moment of $X(t)$ is zero; the latter assumption involves no loss of generality. Before proposing the definition of locally stationary on an interval, we introduce the concept of partitioning of the parameter space I .

Definition 2.9. *Let $I \subset \mathbb{R}$ be an interval (possibly \mathbb{R}). A partition of I is a countable collection of subintervals $\{B_1, B_2, \dots\}$ where $B_k \subset I$ is an interval and k belongs to some countable index set \mathcal{I} , such that*

1. $B_i \cap B_k = \emptyset$ for all $i \neq k$ in \mathcal{I} ,
2. $\cup_{k \in \mathcal{I}} B_k = I$.

In the Silverman sense we have the following definition.

Definition 2.10. *The random process $X(t)$, defined for all real t , is locally stationary in the wide sense, or has a locally stationary covariance, if its covariance can be written as*

$$R_X(t, s) = q(t + s)c(t - s)$$

where $R_X(t, s)$ is the covariance function of $X(t)$ and $q(t + s)$ is the average power of the process at the point $t + s$. The quantity $E[|X(t)|^2]$ is the average instantaneous power of the process $X(t)$. The symmetrization has been chosen because it makes $R_X(t, s)$ Hermitian and because of its suggestive meaning as the average or centroid of the points t and s . We have chosen the correlation function or stationary covariance $c(\tau)$ to be normalized, by which we mean that $c(0) = 1$. The fact that $R_X(t, s)$ is Hermitian implies that $c(\tau) = c^*(-\tau)$.

2.4 Exponentially convex function

No we give a brief description of exponentially convex process [Ehm et. al. \(2003\)](#), [Girardin and Senoussi \(2003\)](#).

Definition 2.11. *The covariance function of a second order zero mean process $\{Z(t), t \in \mathbb{R}\}$ with finite variance, $\text{cov}(Z(t_i), \overline{Z(t_j)}) = \psi(t_i + t_j)$, is called exponentially convex if and only if a complex-valued function C defined on the product space $\mathbb{R} \times \mathbb{R}$ satisfies*

$$\sum_{i=1}^n \sum_{j=1}^n a_i \overline{a_j} \psi(t_i + t_j) \geq 0$$

for all finite sets of complex coefficients a_1, \dots, a_n and points $t_1, \dots, t_n \in \mathbb{R}$.

Example 2.12. *Let $\{Z(t), t \in \mathbb{R}\}$ be a stochastic process with the covariance function $\text{cov}(Z(t), Z(s)) = \psi(t + s)$,*

$$\psi(u) = (1 + u^2)e^{-\frac{u^2}{2}}, \quad u \in \mathbb{R}.$$

As $\psi(u)$ is an exponentially convex function [Ehm et. al. \(2003\)](#) and $\{Z(t), t \in \mathbb{R}\}$ is an exponentially convex process.

We conclude the following theorem from the result of Berg et al. [Berg et. al. \(1984\)](#) (Theorem 5.11, p.12). The corresponding theorem shows that a continuous function is exponentially convex if and only if it is the Laplace transform of a non-negative finite measure.

Theorem 2.13. *Any continuous positive definite function ψ has the representation*

$$\psi(t) = \int_{\mathbb{R}} e^{\lambda t} d\mu(\lambda), \quad t \in \mathbb{R}$$

for a uniquely determined positive Radon measure (a Borel measure that is finite on compact sets) μ .

By Theorem 2.3 and the Karhunen-Loeve theorem, it follows that for any symmetric and continuous, in the mean square, process $Z(\cdot)$ there exists a second order stochastic process F with orthogonal increments and spectral measure μ such that

$$Z(t) = \int_{\mathbb{R}} e^{\lambda t} dF(\lambda), \quad t \in \mathbb{R}.$$

3 Main results: A continuous time PC-LS model

We introduce a new class of non-stationary process which we call certain periodically correlated multi-component locally stationary (PC-LS) and is a combination of two locally stationary and periodically correlated processes as

$$X(t) = X^{ls}(t) + X^p(t), \quad t > 0 \tag{3.1}$$

where $X^{ls}(\cdot)$ is a multi-component LS process and $X^p(\cdot)$ is a sequence of PC process that are independent. We partition the positive real line index space of the process, by disjoint intervals $B_j = (s_{j-1}, s_j]$ where $s_0 = 0$ and $|B_j| = |B_{j+T}|$ for all $j = 1, \dots, T$ and $T \in \mathbb{N}$.

3.1 Continuous PC process

We define M_j as an orthogonally scattered random measure on Borel field of subsets of B_j , by $X_j^p(t) := M_j(s_{j-1}, t]$ for $t \in B_j$ and $X_j^p := M_j(B_j)$, $j \in \mathbb{N}$ where $\{X_j^p(t)\}$ is a positive second order discrete time PC process with period T of centered random variables. For $t \in B_i$, $\{X_i^p(t)\} \stackrel{d}{=} \{X_{kT+i}^p(t)\}$, $i = 1, \dots, T$ and $k = 0, 1, 2, \dots$, in which $S = \sum_{i=1}^T |B_i|$ and $\stackrel{d}{=}$ means equality of the finite dimensional distributions. Thus

$$X^p(t) = \sum_{j=1}^{\infty} X_j^p(t) I_{B_j}(t), \quad t > 0 \tag{3.2}$$

is a continuous time PC process. By the following projection, we provide a linear approximation of the process say the traffic flow in the subintervals of each partition.

Let $A \subset B_j$, $L_A^j = \overline{\text{span}}\{M_j(D), D \subset A\}$ and $\mathcal{P}_A^j : L_{B_j}^j \rightarrow L_A^j$ for $j \in \mathbb{N}$ be an orthogonal projection which is defined by

$$\mathcal{P}_A^j X_j^p = M_j(A). \tag{3.3}$$

with the following property

$$\mathcal{P}_A^j \mathcal{P}_B^j = \mathcal{P}_{A \cap B}^j \tag{3.4}$$

Our idea for providing a proper bilateral correlation of such variables and defining the covariance structure of random measure M_j on subintervals of B_j is the followings. For $A_1, A_2, B \in B_j$ where

$|A_1| = |A_2|$, we assume that

$$\text{corr}(M_j(A_1), M_j(B)) = \text{corr}(M_j(A_2), M_j(B))$$

and by the fact that the flow of traffic on each subinterval of B_j with fixed length have the same multivariate distribution with other random measure as the one of the flow of traffic on any other subintervals of B_j with the same length. For $A \subset B_j$ and $B \subset B_j$ we define the covariance function of M_j as

$$\langle M_j(A), M_j(B) \rangle = \frac{2|A||B|}{a_j(|A| + |B|)} \gamma_{jj}^p \quad (3.5)$$

where $\langle X, Y \rangle = \text{cov}(X, Y) = E[XY]$, $\gamma_{jk}^p = E[X_j^p X_k^p]$ and $|B_j| = a_j$, $j, k \in \mathbb{N}$. The correlation function of this measure is

$$\text{corr}(M_j(A), M_j(B)) = \frac{2\sqrt{|A||B|}}{|A| + |B|} = \frac{2}{\sqrt{\frac{|A|}{|B|} + \frac{|B|}{|A|} + 2}}.$$

As for positive x , $f(x) = x + 1/x \geq 2$ and equality holds for $x = 1$, so the correlation function is equal to one when $|A| = |B|$. Also the correlation function decreases when the amounts of $|A|$ and $|B|$ differ and increases when they approach to each other.

For $A \subset B_j$ and $B \subset B_k$, $j \neq k$

$$\langle M_j(A), M_k(B) \rangle = \frac{|A||B|}{a_j a_k} \gamma_{jk}^p. \quad (3.6)$$

If $B = B_k$ then $\frac{|B|}{a_k} = 1$, so for $j \neq k$

$$\langle M_j(A), M_k(B_k) \rangle = \frac{|A|}{a_j} \gamma_{jk}^p. \quad (3.7)$$

One can easily verify that, this inner product is well defined. In this case for $t, u \in B_j$ and $t \leq u$ we have

$$\langle M_j(s_{j-1}, t], M_j(s_{j-1}, u] \rangle = \frac{2a_j^t a_j^u}{a_j(a_j^t + a_j^u)} \gamma_{jj}^p$$

and for $t \in B_j$, $u \in B_k$

$$\langle M_j(s_{j-1}, t], M_k(s_{k-1}, u] \rangle = \frac{a_j^t a_k^u}{a_j a_k} \gamma_{jk}^p$$

where $a_j^t = t - s_{j-1}$, $a_j = |B_j|$ and $j \in \mathbb{N}$.

3.2 Multi-component LS process

We define a sequence of zero mean stationary process $\{Y_j^s(t), t \in B_j \cup B_{j+1}\}$, where $B_j = (s_{j-1}, s_j]$, $j \in \mathbb{N}$ and $B_0 = \emptyset$ where $Y_j^s(t) \equiv Y_j^s(s_{j-1}, t]$ and $Y_j^s(0) = 0$. We also consider $X_j^{ls}(t)$ as

$$X_j^{ls}(t) = U^{j-1}(t)Y_{j-1}^s(t) + U^j(t)Y_j^s(t) \tag{3.8}$$

for $t \in B_j, j = 2, \dots, T$ and $X_1^{ls}(t) = Y_1^s(t)$ for $t \in B_1$. Also $Y_0^s(\cdot) = 0$ and $\{U^j(\cdot)\}$ is a random weight with exponentially convex covariance which is independent to the processes $\{Y_j^s(\cdot)\}$. We call $X_j^{ls}(t)$, a multi-component locally stationary process as its covariance function is of the form (3.10). Relation (3.8) explains an exponentially convex mixture of two stationary processes. Let

$$X^{ls}(t) = \sum_{j=1}^{\infty} X_j^{ls}(t)I_{B_j}(t), \quad t > 0. \tag{3.9}$$

We show in Theorem 3.1 that $X_j^{ls}(\cdot)$ is a zero mean multi-component LS process indexed by subsets of subintervals of B_j . Thus $X^{ls}(t)$ is a simple multi-component LS process.

3.3 Covariance function of PC-LS process

By the following theorem, we find the covariance structure of the introduced model in this section.

Theorem 3.1. *Let $\{B_j\}$ be the partition of positive real line defined in subsection 3.2. The covariance function of the multi-component PC-LS process $X(t) = X^{ls}(t) + X^p(t)$ where $X^{ls}(t)$ and $X^p(t)$ are independent and defined by (3.9) and (3.2) is $\gamma(t, u) = \gamma^{ls}(t, u) + \gamma^p(t, u)$ where*

$$\gamma^{ls}(t, u) = \sum_{m=1}^{\infty} \gamma_m^{ls}(t, u)I_{B_m}(t) \tag{3.10}$$

and

$$\gamma^p(t, u) = \begin{cases} \sum_{m=1}^{\infty} \frac{2a_m^t a_n^u}{a_m(a_m^t + a_n^u)} \gamma_{mm}^p I_{B_m}(t) & m = n \\ \sum_{m=1}^{\infty} \frac{a_m^t a_n^u}{a_m a_n} \gamma_{mn}^p I_{B_m}(t) & m < n \end{cases}$$

in which

$$\gamma_m^{ls}(t, u) = E[X_m^{ls}(t)X^{ls}(u)] = \sum_{n=m-1}^{m+1} \Gamma_{m,n}(t, u)$$

and $\Gamma_{m,n}(t, u) = E[X_m^{ls}(t)X_n^{ls}(u)], t \in B_m, u \in B_n$.

Proof: According to (3.8) and the fact that $U^j(\cdot)$ and $Y_j^s(\cdot), j \in \mathbb{N}$ are independent processes, for $t, u \in B_m, t \leq u$ we have

$$\Gamma_{m,m}(t, u) = E[X_m^{ls}(t)X_m^{ls}(u)] = \psi_{m-1}(t + u)\gamma_{m-1}(t - u) + \psi_m(t + u)\gamma_m(t - u).$$

For $t \in B_m$ and $u \in B_{m+1}$

$$\Gamma_{m,m+1}(t, u) = E[X_m^{ls}(t)X_{m+1}^{ls}(u)] = \psi_m(t + u)\gamma_m(t - u).$$

For $t \in B_m$ and $u \in B_{m-1}$

$$\Gamma_{m,m-1}(t, u) = E[X_m^{ls}(t)X_{m-1}^{ls}(u)] = \psi_{m-1}(t + u)\gamma_{m-1}(t - u)$$

and for the other cases the process is uncorrelated. Then we have $\gamma_m^{ls}(t, u)$ as the result. This covariance function confirms that, $X_j^{ls}(\cdot)$ is multi-component LS process in the Silverman sense. Therefore

$$\begin{aligned} \gamma^{ls}(t, u) &= E[X^{ls}(t)X^{ls}(u)] = E\left[\sum_{m=1}^{\infty} X_m^{ls}(t)I_{B_m}(t)X^{ls}(u)\right] \\ &= \sum_{m=1}^{\infty} E[X_m^{ls}(t)X^{ls}(u)]I_{B_m}(u) = \sum_{m=1}^{\infty} \gamma_m^{ls}(t, u)I_{B_m}(t). \end{aligned}$$

The covariance function of $X^p(\cdot)$ is $\gamma^p(t, u) = E[X^p(t)X^p(u)]$, for $t \in B_m, u \in B_n, m, n \in \mathbb{N}$ and by (3.2) we have

$$\gamma^p(t, u) = E\left[\left(\sum_{m=1}^{\infty} X_m^p(t)I_{B_m}(t)\right)\left(\sum_{n=1}^{\infty} X_n^p(u)I_{B_n}(u)\right)\right] = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} E[X_m^p(t)X_n^p(u)]I_{B_m}(t)I_{B_n}(u).$$

Thus by (3.5) and (3.6) we have the result.

3.4 Spectral representation

By the integral representation of stationary and exponentially convex processes as mentioned in preliminaries section, the spectral representation of the process $X^{ls}(t)$ is provided. Using the orthogonal projection \mathcal{P}_A^j defined by (3.3) and the spectral representation of discrete time PC process $\{X_j^p\}_{j=1}^{\infty}$, spectral representation of the continuous time PC process $X^p(t)$ is obtained. Finally the harmonizable representation of the multi-component PC-LS process $X(\cdot)$ and its spectral measure is characterized.

Lemma 3.2. *Let $\{X_j^p\}_{j=1}^{\infty}$ be a sequence of PC process with period T . Also $X_j^p = V^j P_j$, where $V = \int_0^{2\pi} e^{i\omega/T} Q(d\omega)$ and P_j is a periodic sequence with period $T, j \in \mathbb{N}$. Then $M_m(s_{m-1}, s_{m-1} + t), 0 < t < |B_m|$, defined by (3.3) can be represented as $\int_0^{2\pi} e^{i\omega m/T} Q(d\omega) \tilde{P}_{m,t}$ in which $\tilde{P}_{m,t}$ is the corresponding periodic sequence to $M_m(s_{m-1}, s_{m-1} + t)$.*

Proof: *By using the result of Proposition 2.2, we can construct a new representation for the random measure M_m as*

$$M_m(s_{m-1}, s_{m-1} + t] = V^m \tilde{P}_{m,t}$$

where V^m is a unitary operator and $\tilde{P}_{m,t}$ is a periodic function in m with period T . Thus by the definition of V^m in Proposition 2.2 we have that for $t \in B_m$

$$M_m(s_{m-1}, s_{m-1} + t) = \int_0^{2\pi} e^{i\omega m/T} Q(d\omega) \tilde{P}_{m,t} = \int_0^{2\pi} e^{i\omega m/T} \zeta_m(d\omega, t)$$

where $0 < t < |B_m|$, $\zeta_m(d\omega, t) = Q(d\omega) \tilde{P}_{m,t}$ is a time varying spectral measure and the covariance function

$$\begin{aligned} \langle M_m(s_{m-1}, t], M_n(s_{n-1}, u) \rangle &= \left\langle \int_0^{2\pi} e^{i\omega m/T} \zeta_m(d\omega, t), \int_0^{2\pi} e^{i\lambda n/T} \zeta_n(d\lambda, u) \right\rangle \\ &= \int_0^{2\pi} \int_0^{2\pi} e^{i\frac{(\omega m - \lambda n)}{T}} \Theta_{m,n}(d\lambda, d\omega, t, u) \end{aligned}$$

where $\Theta_{m,n}(d\lambda, d\omega, t, u) = \langle \zeta_m(d\lambda, t), \zeta_n(d\omega, u) \rangle$.

Theorem 3.3. *The spectral representation of the multi-component PC-LS process $X(t) = X^p(t) + X^{ls}(t)$ is*

$$X(t) = \int_{-\infty}^{\infty} e^{i\lambda t} Z_m(d\lambda, t), \quad t \in B_m \tag{3.11}$$

where the time varying spectral measure $Z_m(\cdot, t)$ is defined as

$$Z_m(d\lambda, t) = \Phi_m(d\lambda, t) + e^{i\lambda(m/T-t)} I_{[0,2\pi)}(d\lambda) \zeta_m(d\lambda, t) \tag{3.12}$$

in which

$$\Phi_m(d\lambda, t) = \phi_{m-1}(d\lambda, t) + \phi_m(d\lambda, t) \tag{3.13}$$

and $\phi_m(d\lambda, t) = U^m(t) \eta_m(d\lambda)$ that $\{U^m(\cdot)\}_{m=1}^{\infty}$ are random variables with exponentially convex covariances. Also

$$F_{m,n}(d\lambda, d\omega, t, u) = \langle \Phi_m(d\lambda, t), \Phi_n(d\omega, u) \rangle = \begin{cases} \psi_{m-1,m-1}(t+u) E[|\eta_{m-1}(d\lambda)|^2] + \psi_{m,m}(t+u) dG_m(\lambda) & n = m, \lambda = \omega \\ \psi_{m,m}(t+u) dG_m(\lambda) & n = m + 1, \lambda = \omega \\ \psi_{m-1,m-1}(t+u) dG_{m-1}(\lambda) & n = m - 1, \lambda = \omega \\ 0 & |n - m| > 1, \lambda \neq \omega \end{cases}$$

$$\Theta_{m,n}(d\lambda, d\omega, t, u) = \langle \zeta_m(d\lambda, t), \zeta_n(d\omega, u) \rangle = \begin{cases} \frac{2a_m^t a_m^u}{a_m(a_m^t + a_m^u)} \tilde{\theta}_{m,m}(d\lambda) & m = n \\ \frac{a_m^t a_n^u}{a_m a_n} \tilde{\theta}_{m,n}(d\lambda) & m \neq n \end{cases}$$

where $\tilde{\theta}_{m,n}(d\lambda) = \langle \xi(d\lambda, m), \xi(d\lambda, n) \rangle$ and $a_m^t = t - s_{m-1}$ for $t \in B_m$, $a_m = |B_m|$. Also $\zeta_m(d\lambda, t) = Q(d\omega) \tilde{P}_{m,t}$. Finally

$$\langle Z_m(d\lambda, t), Z_n(d\omega, u) \rangle = F_{m,n}(d\lambda, d\omega, t, u) + A \Theta_{m,n}(d\lambda, d\omega, t, u)$$

where $A = e^{im/T(\lambda+\omega)-i(\lambda t+\omega u)} I_{[0,2\pi)}(d\lambda) I_{[0,2\pi)}(d\omega)$.

Proof: For the spectral representation of $X^{ls}(t)$, first we find the spectral representation of $X_j^{ls}(t)$, $j \in \mathbb{N}$. By the result of Theorem 2.1, (2.1) and the definition of the process in (3.8), we have

$$X_m^{ls}(t) = U^{m-1}(t) \int_{-\infty}^{\infty} e^{i\lambda t} d\eta_{m-1}(\lambda) + U^m(t) \int_{-\infty}^{\infty} e^{i\lambda t} d\eta_m(\lambda).$$

So $X_m^{ls}(\cdot)$ is a mixture of two harmonizable processes with exponentially convex weights. Also

$$X_m^{ls}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \phi_m(d\lambda, t) + \int_{-\infty}^{\infty} e^{i\lambda t} \phi_{m-1}(d\lambda, t)$$

where $\phi_m(d\lambda, t) = U^m(t)\eta_m(d\lambda)$. Thus

$$X_m^{ls}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \Phi_m(d\lambda, t), \quad t \in B_m \tag{3.14}$$

where $\Phi_m(d\lambda, t)$ is defined by (3.13), so orthogonally scattered property of $\Phi_j(d\lambda, t)$ follows from such property of $\eta_j(\cdot)$ for $j \in \mathbb{N}$. Thus the correlation of the spectral measure $\Phi_m(d\lambda, \cdot)$ is defined as

$$\begin{aligned} F_{m,n}(d\lambda, d\omega, t, u) &= E[\Phi_m(d\lambda, t)\Phi_n(d\omega, u)] \\ &= E\left[(U^{m-1}(t)\eta_{m-1}(d\lambda) + U^m(t)\eta_m(d\lambda))(U^{n-1}(u)\eta_{n-1}(d\omega) + U^n(u)\eta_n(d\omega))\right]. \end{aligned}$$

Which by a simple calculation we obtain $F_{m,n}(d\lambda, d\omega, t, u)$ as expressed by the theorem.

For finding the spectral representation of $X^p(t)$, $t \in \mathbb{R}$ we have

$$X^p(t) = \sum_{m=1}^{\infty} X_m^p(t) I_{B_m}(t) = \sum_{m=1}^{\infty} M_m(s_{m-1}, t] I_{B_m}(t)$$

where M_m is the random measure corresponding to partition B_m .

Then from the assumption in the theorem we have

$$\langle M_m(s_{m-1}, t], M_m(s_{m-1}, u] \rangle = \frac{2a_m^t a_m^u}{a_m(a_m^t + a_m^u)} \int_0^{2\pi} F_{mm}(d\lambda)$$

and in the same way

$$\langle M_m(s_{m-1}, t], M_n(s_{n-1}, u] \rangle = \frac{a_m^t a_n^u}{a_m a_n} \int_0^{2\pi} e^{i\omega \frac{(m-n)}{T}} F_{mn}(d\lambda).$$

In this case, the spectral representation of $X_m^p(t) = M_m(s_{m-1}, t]$ is

$$X_m^p(t) = \int_0^{2\pi} e^{i\lambda m/T} \zeta_m(d\lambda, t), \quad t \in B_m \quad (3.15)$$

According to equations (3.14) and (3.15) we have

$$\begin{aligned} X(t) &= \sum_{m=1}^{\infty} \left(\int_{-\infty}^{\infty} e^{i\lambda t} \Phi_m(d\lambda, t) + \int_0^{2\pi} e^{i\lambda m/T} \zeta_m(d\lambda, t) \right) \\ &= \sum_{m=1}^{\infty} \left(\int_{-\infty}^{\infty} e^{i\lambda t} \Phi_m(d\lambda, t) + \int_0^{2\pi} e^{i\lambda t} e^{i\lambda(m/T-t)} \zeta_m(d\lambda, t) \right) \\ &= \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} e^{i\lambda t} \left(\Phi_m(d\lambda, t) + e^{i\lambda(m-t)} I_{[0,2\pi)}(d\lambda) \right). \end{aligned}$$

Thus the aggregated process has a time varying spectral representation by (3.11).

References

- C. Berg, J.P. Christensen, P. Ressel, "Harmonic Analysis on Semigroups", New York: Springer-Verlag, 1984.
- S.M. Berman, "Local nondeterminism and local times of Gaussian processes", *Indiana Univ. Math. J.* 23, pp. 6994, 1974.
- D.R. Cox, H.D. Miller, "The Theory of Stochastic Processes", Chapman and Hall, 1994.
- D. Dehay, H.L. Hurd, "Representation and Estimation for Periodically and Almost Periodically Correlated Random Processes", in: *W.A. Gardner (Ed), Cyclostationarity in Communications and Signal Processing, IEEE Press, New York*, pp. 295-326, 1994.
- D. Dehay, A. Loughani, "Locally Harmonizable Covariance: Spectral Analysis", *Kybernetika*, Vol.30, No.5, pp.543-550, 1994.
- W. Ehm, M.G. Genton, T. Gneiting, "Stationary covariances associated with exponentially convex functions", *Bernoulli*, No. 9(4), pp.607-615, 2003.
- W.A. Gardner, "Cyclostationarity in Communications and Signal Processing", *IEEE Press, New York*, 1994.

- V. Girardin, R. Senoussi, "Semigroup stationary processes and spectral representation", *Bernoulli*, No. 9(5), pp.857-876, 2003.
- E.G. Gladyshev, "Periodically Correlated Random Sequences", *Soviet Math. Dokl.*, No.2, pp.385-388, 1961.
- H.L. Hurd, A.G. Miamee, "Periodically Correlated Random Sequences: Spectral Theory and Practice", John Wiley, 2007.
- M. Loeve, "Fonctions aleatoires a caracte're exponentiel" In P. Levy, *Processus Stochastiques et Mouvement Brownien*. 2nd edn., pp. 409-420. Paris: Gauthiers-Villars, 1965.
- M.E. Oxley, T.F. Reid, B.W. Suter, "Locally stationary processes", *IEEE Workshop on statistical signal and array process*, 2000.
- R.A. Silverman, "Locally Stationary Random Processes", *IRE Trans. Inf. Theory*, Vol.IT-3, No.3, pp.182-187, 1957.
- P. Wahlberg, M. Hansson, "Kernels and Multiple Windows for Estimation of the Wigner-Ville Spectrum of Gaussian Locally Stationary Processes", *IEEE Transactions on signal processing*, Vol.55, No.1, 2007.



Prediction via bivariate concomitants of bivariate order statistics

Seyyed Roohollah Roozegar¹, Ahad Jamalizadeh², Mehdi amiri³

¹Department of Mathematics, Yasouj University, Yasouj, Iran

²Department of Statistics, Faculty of Mathematics & Computer, Shahid Bahonar
University of Kerman, Kerman, Iran

³Department of Statistics, Homozgan University, Bandar Abbas, Iran

Abstract: In this paper, by considering a $(2n)$ -dimensional elliptically random vector $(\mathbf{X}^T, \mathbf{Y}^T)^T$, we derive the joint distribution of bivariate concomitants of order statistics $(X_{[r]}, Y_{[s]})^T$. Further, by considering an elliptical distribution for the $(2n + 1)$ -dimensional random vector $(X_0, \mathbf{X}^T, \mathbf{Y}^T)^T$, and treating X_0 as a covariate variable, we present mixture representations for the joint distributions $(X_0, X_{[r]}, Y_{[s]})^T$ and $(X_0, aX_{[r]} + bY_{[s]})^T$. These mixture representations enables us to obtain the best (nonlinear) predictors of X_0 based on $X_{[r]}$ and $Y_{[s]}$, and X_0 based on $aX_{[r]} + bY_{[s]}$.

Keywords: Bivariate Concomitants, Elliptical distribution, Mixture distribution, Multivariate unified skew-elliptical distribution, Order statistics.

Mathematics Subject Classification (2010): 62E10 62E15 62H10.

1 Introduction

Let (X_i, Y_i) , $i = 1, \dots, n$ be a random sample from an absolutely continuous bivariate distribution, distributed commonly as $(X, Y)^T$. Order the sample by X -variate, and obtain the order statistics $X_{(1)}, \dots, X_{(n)}$, then the Y -variate associated with the s th order statistics $X_{(s)}$ is called the concomitant of the s th order statistic, and is conventionally denoted by $Y_{[s]}$. [David \(1973\)](#) and [Bhattacharya \(1974\)](#) discuss various distributional properties and inferential result concerning to univariate concomitants of order statistics. [Wang and Nagaraja \(2009\)](#) discuss the properties of concomitants for dependent samples. Recently [Jamalizadeh and Balakrishnan \(2012\)](#) consider the more general structure. They have been considered $\mathbf{X} = (X_1, \dots, X_n)^T$ and $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ be two random vectors of dimension n following the elliptical distribution as

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim EC_{2n} \left(\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{pmatrix}, h^{(2n)} \right), \quad (1.1)$$

¹Speaker's Name : Ehsanroozegar@yahoo.com

then they have been derived the exact distribution of univariate concomitants of order statistics. In this paper we develop the result of univariate concomitants to bivariate concomitants of order statistics. Specifically, by assuming the distribution in (1.1). For describe the problem briefly, consider a survey involving n students in a university. Let (X_i, Y_i) , $i = 1, \dots, n$ contain the scores of the i th-student in midterm and final exams, respectively. Further, let $X_{(1)} < \dots < X_{(n)}$ and $Y_{(1)} < \dots < Y_{(n)}$, be the vectors of order statistics corresponding to \mathbf{X} and \mathbf{Y} , respectively. The best students in midterm and final exams, say $X_{(n)}$ and $Y_{(n)}$ are of interest, but the corresponding concomitants $X_{[n]}$ (the midterm score of the best student in final exam) and $Y_{[n]}$ (the final score of the best student in midterm exam) are of interest as well. Note that in this case the students are in a class and have been considered under an educational program, and so the assumption of independence is not realistic.

The paper is organized as follows. In section 2, we first provide a short introduction to multivariate unified skew-elliptical distributions. In section 3, we present a mixture representation for distribution of $(X_{[r]}, Y_{[s]})^T$. In section 4, by considering an elliptical distribution for the $(2n + 1)$ -dimensional random vector $(X_0, \mathbf{X}^T, \mathbf{Y}^T)^T$, and treating X_0 as a covariate variable, we have presented mixture representation for distributions $(X_0, X_{[r]}, Y_{[s]})^T$ and $(X_0, aX_{[r]} + bY_{[s]})^T$. These mixture representations enables us to obtain the best (nonlinear) predictors X_0 based on $X_{[r]}$ and $Y_{[s]}$, and X_0 based on $aX_{[r]} + bY_{[s]}$, and so on.

2 Preliminaries

The following notations will be used throughout this paper: $\Phi_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the cdf of $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $T_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ for the cdf of the n -variate t distribution, $F_{EC_n}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)})$ and $f_{EC_n}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)})$ for the cdf and pdf of the n -variate elliptical distribution with location parameter $\boldsymbol{\mu}$, dispersion matrix $\boldsymbol{\Sigma}$, and density generator function $h^{(n)}$, respectively: $EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)})$, and simply $\Phi_n(\cdot; \boldsymbol{\Sigma})$, $T_n(\cdot; \boldsymbol{\Sigma}, \nu)$ and $F_{EC_n}(\cdot; \boldsymbol{\Sigma}, h^{(n)})$ for the case $\boldsymbol{\mu} = \mathbf{0}$.

In this section we give a brief review of multivariate unified skew-elliptical (SUE) distribution presented by [Arellano-Valle and Azzalini \(2006\)](#).

Let \mathbf{U} and \mathbf{V} be two random vectors of dimensions m and n , respectively, such that

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim EC_{m+n} \left(\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Omega} \end{pmatrix}, h^{(m+n)} \right), \quad (2.1)$$

then, the n -dimensional random vector \mathbf{X} is said to have the SUE distribution with parameter $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$ and density generator function $h^{(m+n)}$, denoted by $\mathbf{X} \sim SUE_{n,m}(\boldsymbol{\theta}, h^{(m+n)})$, if $\mathbf{X} \stackrel{d}{=} \mathbf{V} \mid (\mathbf{U} > \mathbf{0})$. The density function of \mathbf{X} is

$$g_{SUE_{n,m}}(\mathbf{x}; \boldsymbol{\theta}, h^{(m+n)})$$

$$= \frac{f_{EC_n}(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(n)})}{F_{EC_m}(\boldsymbol{\eta}; \boldsymbol{\Gamma}, h^{(m)})} \times F_{EC_m}\left(\boldsymbol{\eta} + \boldsymbol{\Lambda}^T \boldsymbol{\Omega}^{-1}(\mathbf{x} - \boldsymbol{\xi}); \boldsymbol{\Gamma} - \boldsymbol{\Lambda}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda}, h_{w(\mathbf{x})}^{(m)}\right), \quad (2.2)$$

for $\mathbf{x} \in \mathbb{R}^n$, where $w(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1}(\mathbf{x} - \boldsymbol{\xi})$ and $h_b^{(m)}(u) = h^{(m+n)}(u + b) / h^{(n)}(b)$ for $u, b > 0$.

Note that for $h^{(m+n)}(u) = (2\pi)^{-(m+n)/2} \exp(-u/2)$ ($u \geq 0$), the multivariate unified skew-normal (SUN) distribution $SUN_{n,m}(\boldsymbol{\theta})$ and for $h^{(m+n)}(u) = \frac{\Gamma(\frac{\nu+m+n}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{\nu\pi}{2})} (1 + \frac{u}{\nu})^{-(\nu+m+n)/2}$, the multivariate unified skew-t (SUT) distribution $SUT_{n,m}(\boldsymbol{\theta}, \nu)$ will be obtained.

Jamalizadeh and Balakrishnan (2012) presented the marginal and conditional distributions of the SUE distributions. This result is given in the following lemma.

Let X_1 and X_2 be two random vectors of dimensions n_1 and $n - n_1$, respectively, where

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim SUE_{n,m}(\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}, h^{(m+n)}).$$

Consider the following partitions of $\boldsymbol{\xi}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$ with the dimensions matching suitably:

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_2 \end{pmatrix}.$$

Then we have the following lemma.

Lemma 2.1. (Jamalizadeh and Balakrishnan (2012))

- (i) $\mathbf{X}_1 \sim SUE_{n_1,m}(\boldsymbol{\xi}_1, \boldsymbol{\eta}, \boldsymbol{\Omega}_{11}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}_1, h^{(m+n_1)})$;
- (ii) $\mathbf{X}_2 | (\mathbf{X}_1 = \mathbf{x}_1) \sim SUE_{n-n_1,m}(\boldsymbol{\xi}^{2.1}(\mathbf{x}_1), \boldsymbol{\eta}^{2.1}(\mathbf{x}_1), \boldsymbol{\Omega}^{22.1}, \boldsymbol{\Gamma}^{2.1}, \boldsymbol{\Lambda}^{2.1}, h_{q(\mathbf{x}_1)}^{(m+n-n_1)})$, where

$$\begin{aligned} \boldsymbol{\xi}^{2.1}(\mathbf{x}_1) &= \boldsymbol{\xi}_2 + \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\xi}_1), & \boldsymbol{\eta}^{2.1}(\mathbf{x}_1) &= \boldsymbol{\eta} + \boldsymbol{\Lambda}_1^T \boldsymbol{\Omega}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\xi}_1), \\ \boldsymbol{\Omega}^{22.1} &= \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12}, & \boldsymbol{\Gamma}^{2.1} &= \boldsymbol{\Gamma} - \boldsymbol{\Lambda}_1^T \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Lambda}_1, \\ \boldsymbol{\Lambda}^{2.1} &= \boldsymbol{\Lambda}_2 - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Lambda}_1, & q_1(\mathbf{x}_1) &= (\mathbf{x}_1 - \boldsymbol{\xi}_1)^T \boldsymbol{\Omega}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\xi}_1). \end{aligned}$$

In the normal and t cases when $n = 1$ whatever the value of $m = 1, 2, \dots$ we can obtain the corresponding moments. For example, if $X \sim SUN_{1,m}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\eta}, \omega, \boldsymbol{\Gamma}, \boldsymbol{\lambda}^T)$, $\boldsymbol{\xi} \in \mathbb{R}$, $\boldsymbol{\eta} \in \mathbb{R}^m$, $\boldsymbol{\Gamma} \in \mathbb{R}^{m \times m}$ is a positive definite dispersion matrix, $\omega > 0$, $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^T$, we have [see Jamalizadeh and Balakrishnan (2012)]

$$E(X) = \boldsymbol{\xi} + \frac{1}{\Phi_m(\boldsymbol{\eta}; \boldsymbol{\Gamma})} \sum_{i=1}^m \frac{\lambda_i}{\sqrt{\gamma_{ii}}} \phi\left(\frac{\eta_i}{\sqrt{\gamma_{ii}}}\right) \Phi_{m-1}\left(\boldsymbol{\eta}_{-i} - \frac{\eta_i}{\gamma_{ii}} \boldsymbol{\gamma}_{-ii}; \boldsymbol{\Gamma}_{-i|i}\right), \quad (2.3)$$

where, for some $i = 1, 2, \dots, m$,

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_{ii} & \boldsymbol{\gamma}_{-ii}^T \\ \boldsymbol{\gamma}_{-ii} & \boldsymbol{\Gamma}_{-i-i} \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} \lambda_i \\ \boldsymbol{\lambda}_{-i} \end{pmatrix}, \quad \begin{pmatrix} \eta_i \\ \boldsymbol{\eta}_{-i} \end{pmatrix},$$

with

$$\boldsymbol{\Gamma}_{-i|i} = \boldsymbol{\Gamma}_{-i-i} - \frac{\boldsymbol{\gamma}_{-ii}\boldsymbol{\gamma}_{-ii}^T}{\gamma_{ii}}. \tag{2.4}$$

Similarly if $X \sim SUT_{1,m}(\boldsymbol{\theta}, \nu)$, then for $\nu > 1$,

$$\begin{aligned} E(X) &= \xi + \frac{\Gamma(\frac{\nu-1}{2})(\nu)^{\nu/2}}{2\sqrt{\pi}\Gamma(\frac{\nu}{2})T_m(\boldsymbol{\eta}; \boldsymbol{\Gamma}, \nu)} \sum_{i=1}^m \frac{\lambda_i}{\sqrt{\gamma_{ii}}} \left(\nu + \frac{\eta_i^2}{\gamma_{ii}}\right)^{-(\nu-1)/2} \\ &\quad \times T_{m-1}\left(\frac{\sqrt{\nu-1}}{\sqrt{\nu + \frac{\eta_i^2}{\gamma_{ii}}}} \left(\boldsymbol{\eta}_{-i} - \frac{\eta_i}{\gamma_{ii}}\boldsymbol{\gamma}_{-ii}\right); \boldsymbol{\Gamma}_{-i|i}, \nu - 1\right). \end{aligned} \tag{2.5}$$

3 Joint distribution of bivariate concomitants

Let $\mathbf{X} = (X_1, \dots, X_n)^T$ and $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ be two random vectors of dimensions n , jointly distributed as in (??) Also, let $\mathbf{X}_{(n)} = (X_{(1)}, \dots, X_{(n)})^T$ and $\mathbf{Y}_{(n)} = (Y_{(1)}, \dots, Y_{(n)})^T$, denote the vectors of order statistics arising from \mathbf{X} and \mathbf{Y} , respectively. Also let $\mathbf{X}_{[n]} = (X_{[1]}, \dots, X_{[n]})^T$ and $\mathbf{Y}_{[n]} = (Y_{[1]}, \dots, Y_{[n]})^T$ be the vectors of concomitants corresponding to $\mathbf{X}_{(n)}$ and $\mathbf{Y}_{(n)}$, respectively. In this section we present a mixture representation for the joint distribution of $(X_{[r]}, Y_{[s]})^T$, for $r = 1, 2, \dots, n$ and $s = 1, 2, \dots, n$, in the terms of multivariate unified skew-elliptical distributions. For this purpose, let, for $i = 1, \dots, n$ and $j = 1, 2, \dots, n$, \mathbf{X} and \mathbf{Y} be partitioned as $\mathbf{X} = (X_i, \mathbf{X}_{-i}^T)^T$ and $\mathbf{Y} = (Y_j, \mathbf{Y}_{-j}^T)^T$ and introduce the following partitions:

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{X}} &= \begin{pmatrix} \mu_{X_i} \\ \boldsymbol{\mu}_{\mathbf{X}_{-i}} \end{pmatrix}, \quad \boldsymbol{\mu}_{\mathbf{Y}} = \begin{pmatrix} \mu_{Y_j} \\ \boldsymbol{\mu}_{\mathbf{Y}_{-j}} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} = \begin{pmatrix} \sigma_{X_i X_i} & \boldsymbol{\sigma}_{\mathbf{X}_{-i} X_i}^T \\ \boldsymbol{\sigma}_{\mathbf{X}_{-i} X_i} & \boldsymbol{\Sigma}_{\mathbf{X}_{-i} \mathbf{X}_{-i}} \end{pmatrix}, \\ \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} &= \begin{pmatrix} \sigma_{Y_j Y_j} & \boldsymbol{\sigma}_{\mathbf{Y}_{-j} Y_j}^T \\ \boldsymbol{\sigma}_{\mathbf{Y}_{-j} Y_j} & \boldsymbol{\Sigma}_{\mathbf{Y}_{-j} \mathbf{Y}_{-j}} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} = \begin{pmatrix} \sigma_{X_i Y_j} & \boldsymbol{\sigma}_{\mathbf{Y}_{-j} X_i}^T \\ \boldsymbol{\sigma}_{\mathbf{X}_{-i} Y_j} & \boldsymbol{\Sigma}_{\mathbf{X}_{-i} \mathbf{Y}_{-j}} \end{pmatrix}. \end{aligned}$$

Also, let n and $1 \leq r \leq n$ be two integers, and for integers $1 \leq j_1 < \dots < j_{r-1} \leq n - 1$, let the diagonal matrix $\mathbf{S}_{j_1 \dots j_{r-1}} = \text{diag}(s_1, \dots, s_{n-1})$ is such that $s_i = \begin{cases} 1, & \text{for } i = j_1, \dots, j_{r-1} \\ -1, & \text{otherwise} \end{cases}$. In particular, $\mathbf{S}_{j_1 \dots j_{n-1}} = \mathbf{I}_{n-1}$ and $\mathbf{S}_{j_0} = -\mathbf{I}_{n-1}$.

Theorem 3.1. *The joint cdf of $(X_{[r]}, Y_{[s]})^T$, for $\mathbf{t} = (t_1, t_2)^T \in \mathbb{R}^2$ and for $r = 1, 2, \dots, n$ and $s = 1, 2, \dots, n$ are the mixtures*

$$F_{[r],[s]}(\mathbf{t}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(2n)}) = \sum_{i=1}^n \sum_{j=1}^n \sum_{\mathbf{i} \in \mathbf{I}} \sum_{\mathbf{j} \in \mathbf{J}} \pi_{ij, \mathbf{i}, \mathbf{j}} G_{SUE_{2,2n-2}}(\mathbf{t}; \boldsymbol{\theta}_{ij, \mathbf{i}, \mathbf{j}}, h^{(2n)}), \tag{3.1}$$

where $G_{SUE_{2,2n-2}}(\cdot; \boldsymbol{\theta}_{ij, \mathbf{i}, \mathbf{j}}, h^{(2n)})$ is cdf of $SUE_{2,2n-2}(\boldsymbol{\theta}_{ij, \mathbf{i}, \mathbf{j}}, h^{(2n)})$ and $\mathbf{I} = \{\mathbf{i} : \mathbf{i} = (i_1 \cdots i_{r-1}), 1 \leq i_1 < \cdots < i_{r-1} \leq n-1\}$, $\mathbf{J} = \{\mathbf{j} : \mathbf{j} = (j_1 \cdots j_{s-1}), 1 \leq j_1 < \cdots < j_{s-1} \leq n-1\}$ and the mixing probabilities are $\pi_{ij, \mathbf{i}, \mathbf{j}} = F_{EC_{2n-2}}(\boldsymbol{\eta}_{ij, \mathbf{i}, \mathbf{j}}; \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}, h^{(2n-2)})$ with $\boldsymbol{\theta}_{ij, \mathbf{i}, \mathbf{j}} = (\boldsymbol{\xi}_{ij}, \boldsymbol{\eta}_{ij, \mathbf{i}, \mathbf{j}}, \boldsymbol{\Omega}_{ij}, \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}, \boldsymbol{\Lambda}_{ij, \mathbf{i}, \mathbf{j}})$ where

$$\begin{aligned} \boldsymbol{\xi}_{ij} &= \begin{pmatrix} \mu_{X_i} \\ \mu_{Y_j} \end{pmatrix}^T, \quad \boldsymbol{\eta}_{ij, \mathbf{i}, \mathbf{j}} = \begin{pmatrix} \mathbf{S}_i (\mu_{Y_i} \mathbf{1}_{n-1} - \boldsymbol{\mu}_{\mathbf{Y}_{-i}}) \\ \mathbf{S}_j (\mu_{X_j} \mathbf{1}_{n-1} - \boldsymbol{\mu}_{\mathbf{X}_{-j}}) \end{pmatrix}, \\ \boldsymbol{\Omega}_{ij} &= \begin{pmatrix} \sigma_{X_i X_i} & \sigma_{X_i Y_j} \\ \sigma_{Y_j Y_j} & \sigma_{Y_j X_i} \end{pmatrix}, \quad \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}} = \begin{pmatrix} \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}^{11} & \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}^{12} \\ \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}^{22} & \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}^{21} \end{pmatrix}, \\ \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}^{11} &= \mathbf{S}_i (\sigma_{Y_i Y_i} \mathbf{1}_{n-1} \mathbf{1}_{n-1}^T - \mathbf{1}_{n-1} \boldsymbol{\sigma}_{\mathbf{Y}_{-i} Y_i}^T - \boldsymbol{\sigma}_{\mathbf{Y}_{-i} Y_i} \mathbf{1}_{n-1}^T + \boldsymbol{\Sigma}_{\mathbf{Y}_{-i} \mathbf{Y}_{-i}}) \mathbf{S}_i, \\ \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}^{12} &= \mathbf{S}_i (\sigma_{X_j Y_i} \mathbf{1}_{n-1} \mathbf{1}_{n-1}^T - \mathbf{1}_{n-1} \boldsymbol{\sigma}_{\mathbf{Y}_{-i} X_j}^T - \boldsymbol{\sigma}_{\mathbf{X}_{-j} Y_i} \mathbf{1}_{n-1}^T + \boldsymbol{\Sigma}_{\mathbf{X}_{-j} \mathbf{Y}_{-i}}) \mathbf{S}_j, \\ \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}^{22} &= \mathbf{S}_j (\sigma_{X_j X_j} \mathbf{1}_{n-1} \mathbf{1}_{n-1}^T - \mathbf{1}_{n-1} \boldsymbol{\sigma}_{\mathbf{X}_{-j} X_j}^T - \boldsymbol{\sigma}_{\mathbf{X}_{-j} X_j} \mathbf{1}_{n-1}^T + \boldsymbol{\Sigma}_{\mathbf{X}_{-j} \mathbf{X}_{-j}}) \mathbf{S}_j, \\ \boldsymbol{\Lambda}_{ij, \mathbf{i}, \mathbf{j}} &= \begin{pmatrix} [\mathbf{S}_i (\sigma_{Y_i X_i} \mathbf{1}_{n-1} - \boldsymbol{\sigma}_{\mathbf{Y}_{-i} X_i})]^T & [\mathbf{S}_j (\sigma_{X_j X_i} \mathbf{1}_{n-1} - \boldsymbol{\sigma}_{\mathbf{X}_{-j} X_i})]^T \\ [\mathbf{S}_i (\sigma_{Y_i Y_j} \mathbf{1}_{n-1} - \boldsymbol{\sigma}_{\mathbf{Y}_{-i} Y_j})]^T & [\mathbf{S}_j (\sigma_{X_j Y_j} \mathbf{1}_{n-1} - \boldsymbol{\sigma}_{\mathbf{X}_{-j} Y_j})]^T \end{pmatrix}. \end{aligned}$$

Proof. we have

$$F_{[r],[s]}(\mathbf{t}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(2n)}) = \sum_{i=1}^n \sum_{j=1}^n \Pr(X_i \leq t_1, Y_j \leq t_2, Y_{(r)} = Y_i, X_{(s)} = X_j). \tag{3.2}$$

Consider (i, j) -th term on the RHS of (3.2), and write it as

$$\begin{aligned} &\Pr(X_i \leq t_1, Y_j \leq t_2, Y_{(r)} = Y_i, X_{(s)} = X_j) \\ &= \sum_{\mathbf{i} \in \mathbf{I}} \sum_{\mathbf{j} \in \mathbf{J}} \Pr(\mathbf{S}_i (\mathbf{1}_{n-1} Y_i - \mathbf{Y}_{-i}) > \mathbf{0}, \mathbf{S}_j (\mathbf{1}_{n-1} X_j - \mathbf{X}_{-j}) > \mathbf{0}) \\ &\quad \times \Pr(X_i \leq t_1, Y_j \leq t_2 | \mathbf{S}_i (\mathbf{1}_{n-1} Y_i - \mathbf{Y}_{-i}) > \mathbf{0}, \mathbf{S}_j (\mathbf{1}_{n-1} X_j - \mathbf{X}_{-j}) > \mathbf{0}) \end{aligned}$$

which readily yields the mixture representation for the cdf. □

Remark 3.2. *By using the result in Theorem 3.1, and the marginal and conditional distributions of the multivariate unified skew-elliptical distribution presented in Lemma 2.1, we can easily find the marginal*

and conditional distributions of the vector $(X_{[r]}, Y_{[s]})^T$. For example, the conditional distribution of $Y_{[s]}$, given $(X_{[r]} = t_1)$ enables us, by the results in (2.3) and (2.5), to obtain the conditional mean $E[Y_{[s]} | (X_{[r]} = t_1)]$ in the normal and t cases. In fact, these conditional means are best (nonlinear) predictors of $Y_{[s]}$ based on $X_{[r]}$, under the squared-error loss.

4 Prediction via bivariate concomitant

In this section we consider the prediction based on bivariate concomitants of order statistics. For this purpose, Let $X_0, \mathbf{X} = (X_1, \dots, X_n)^T$ and $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ jointly distributed as

$$\begin{pmatrix} X_0 \\ \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim EC_{2n+1} \left(\boldsymbol{\mu}^* = \begin{pmatrix} \mu_{X_0} \\ \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{pmatrix}, \boldsymbol{\Sigma}^* = \begin{pmatrix} \sigma_{X_0 X_0} & \boldsymbol{\sigma}_{\mathbf{X} X_0}^T & \boldsymbol{\sigma}_{\mathbf{Y} X_0}^T \\ & \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \\ & & \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{Y}} \end{pmatrix}, h^{(2n+1)} \right). \tag{4.1}$$

Then, like as Theorem ??, we can show that the distribution of $(X_0, X_{[r]}, Y_{[s]})^T$ and $(X_0, aX_{[r]} + bY_{[s]})^T$ where $a \in \mathbb{R}, b \in \mathbb{R}$ can be expressed as a mixture of the SUE distributions. For this purpose, we need the following partitions, for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, n$,

$$\boldsymbol{\sigma}_{\mathbf{X} X_0} = \begin{pmatrix} \sigma_{X_j X_0} \\ \boldsymbol{\sigma}_{\mathbf{X}_{-j} X_0} \end{pmatrix}, \boldsymbol{\sigma}_{\mathbf{Y} X_0} = \begin{pmatrix} \sigma_{Y_i X_0} \\ \boldsymbol{\sigma}_{\mathbf{Y}_{-i} X_0} \end{pmatrix}.$$

Theorem 4.1. (i) The joint cdf of $(X_0, X_{[r]}, Y_{[s]})^T$, for $\mathbf{t} = (t_1, t_2, t_3)^T \in \mathbb{R}^3$, are the mixtures

$$F_{(X_0, X_{[r]}, Y_{[s]})}(\mathbf{t}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, h^{(2n+1)}) = \sum_{i=1}^n \sum_{j=1}^n \sum_{\mathbf{i} \in \mathbf{I}} \sum_{\mathbf{j} \in \mathbf{J}} \pi_{ij, \mathbf{i}, \mathbf{j}} \times G_{SUE_{3, 2n-2}}(\mathbf{t}; \boldsymbol{\theta}_{ij, \mathbf{i}, \mathbf{j}}^*, h^{(2n+1)}),$$

(ii) The joint cdf of $(X_0, aX_{[r]} + bY_{[s]})^T$, for $\mathbf{t} = (t_1, t_2)^T \in \mathbb{R}^2$, are the mixtures

$$F_{(X_0, aX_{[r]} + bY_{[s]})}(\mathbf{t}; \boldsymbol{\mu}', \boldsymbol{\Sigma}', h^{(2n+1)}) = \sum_{i=1}^n \sum_{j=1}^n \sum_{\mathbf{i} \in \mathbf{I}} \sum_{\mathbf{j} \in \mathbf{J}} \pi_{ij, \mathbf{i}, \mathbf{j}} \times G_{SUE_{2, 2n-2}}(\mathbf{t}; \boldsymbol{\theta}'_{ij, \mathbf{i}, \mathbf{j}}, h^{(2n+1)}),$$

where $\pi_{ij, \mathbf{i}, \mathbf{j}}, \boldsymbol{\eta}_{ij, \mathbf{i}, \mathbf{j}}$ and $\boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}$ are as in Theorem 3.1, $\boldsymbol{\theta}_{ij, \mathbf{i}, \mathbf{j}}^* = (\boldsymbol{\xi}_{ij}^*, \boldsymbol{\eta}_{ij, \mathbf{i}, \mathbf{j}}, \boldsymbol{\Omega}_{ij}^*, \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}, \boldsymbol{\Lambda}_{ij, \mathbf{i}, \mathbf{j}}^*)$ and $\boldsymbol{\theta}'_{ij, \mathbf{i}, \mathbf{j}} = (\boldsymbol{\xi}'_{ij}, \boldsymbol{\eta}_{ij, \mathbf{i}, \mathbf{j}}, \boldsymbol{\Omega}'_{ij}, \boldsymbol{\Gamma}_{ij, \mathbf{i}, \mathbf{j}}, \boldsymbol{\Lambda}'_{ij, \mathbf{i}, \mathbf{j}})$ where

$$\boldsymbol{\xi}_{ij}^* = \begin{pmatrix} \mu_{X_0} \\ \mu_{X_i} \\ \mu_{Y_j} \end{pmatrix}, \boldsymbol{\Omega}_{ij}^* = \begin{pmatrix} \sigma_{X_0 X_0} & \sigma_{X_i X_0} & \sigma_{Y_j X_0} \\ & \sigma_{X_i X_i} & \sigma_{X_i Y_j} \\ & & \sigma_{Y_j Y_j} \end{pmatrix}, \boldsymbol{\xi}'_{ij} = \begin{pmatrix} \mu_{X_0} \\ a\mu_{X_i} + b\mu_{Y_j} \end{pmatrix},$$

$$\Lambda_{ij,i,j}^* = \begin{pmatrix} [\mathbf{S}_i (\sigma_{Y_i X_0} \mathbf{1}_{n-1} - \sigma_{\mathbf{Y}_{-i} X_0})]^T & [\mathbf{S}_j (\sigma_{X_j X_0} \mathbf{1}_{n-1} - \sigma_{\mathbf{X}_{-j} X_0})]^T \\ [\mathbf{S}_i (\sigma_{Y_i X_i} \mathbf{1}_{n-1} - \sigma_{\mathbf{Y}_{-i} X_i})]^T & [\mathbf{S}_j (\sigma_{X_j X_i} \mathbf{1}_{n-1} - \sigma_{\mathbf{X}_{-j} X_i})]^T \\ [\mathbf{S}_i (\sigma_{Y_i Y_j} \mathbf{1}_{n-1} - \sigma_{\mathbf{Y}_{-i} Y_j})]^T & [\mathbf{S}_j (\sigma_{X_j Y_j} \mathbf{1}_{n-1} - \sigma_{\mathbf{X}_{-j} Y_j})]^T \end{pmatrix},$$

$$\Omega'_{ij} = \begin{pmatrix} \sigma_{X_0 X_0} & a\sigma_{X_i X_0} + b\mu_{Y_j} \\ a^2\sigma_{X_i X_i} + b^2\sigma_{Y_j Y_j} + 2ab\sigma_{X_i Y_j} \end{pmatrix}, \Lambda'_{ij,i,j} = \begin{pmatrix} \Lambda'_{ij,i,j}{}^{11} & \Lambda'_{ij,i,j}{}^{12} \\ \Lambda'_{ij,i,j}{}^{21} & \Lambda'_{ij,i,j}{}^{22} \end{pmatrix},$$

$$\Lambda'_{ij,i,j}{}^{11} = [\mathbf{S}_i (\sigma_{Y_i X_0} \mathbf{1}_{n-1} - \sigma_{\mathbf{Y}_{-i} X_0})]^T, \Lambda'_{ij,i,j}{}^{12} = [\mathbf{S}_j (\sigma_{X_j X_0} \mathbf{1}_{n-1} - \sigma_{\mathbf{X}_{-j} X_0})]^T,$$

$$\Lambda'_{ij,i,j}{}^{21} = [\mathbf{S}_i ((a\sigma_{Y_i X_i} + b\sigma_{Y_i Y_j}) \mathbf{1}_{n-1} - a\sigma_{\mathbf{Y}_{-i} X_i} - b\sigma_{\mathbf{Y}_{-i} Y_j})]^T,$$

$$\Lambda'_{ij,i,j}{}^{22} = [\mathbf{S}_j ((a\sigma_{X_j X_i} + b\sigma_{X_j Y_j}) \mathbf{1}_{n-1} - a\sigma_{\mathbf{X}_{-j} X_i} - b\sigma_{\mathbf{X}_{-j} Y_j})]^T.$$

Remark 4.2. By using the result in Theorem 4.1, and the marginal and conditional distributions of the SUE distributions presented in Lemma 2.1, we can find the marginal and conditional distributions of the vector $(X_0, X_{[r]}, Y_{[s]})^T$ and $(X_0, aX_{[r]} + bY_{[s]})$ for $r = 1, \dots, n$ and $s = 1, \dots, n$. For example, the conditional distribution of X_0 , given $(X_{[n]} = t_2, Y_{[n]} = t_3)$ and $(X_{[n]} + Y_{[n]} = t)$ enables us, by the results in (2.3) and (2.5), to obtain the conditional mean $E[X_0 | (X_{[n]} = t_2, Y_{[n]} = t_3)]$ and $E[X_0 | (X_{[n]} + Y_{[n]} = t)]$ in the normal and t cases, respectively.

Example 4.3. Suppose that $(X_1, X_2, Y_1, Y_2) \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (1.65, 1.65, 3.5, 3.5)^T$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.044 & 0.044 & 0.080 & 0.085 \\ 0.044 & 0.047 & 0.087 & 0.096 \\ 0.080 & 0.087 & 0.297 & 0.300 \\ 0.085 & 0.096 & 0.300 & 0.343 \end{bmatrix}.$$

Let $G_{SUN_{2,2}}(\cdot; \boldsymbol{\theta})$ denote the cdf of $SUN_{2,2}(\boldsymbol{\theta})$. So from Theorem ?? we have $F_{[2],[2]}((t_1, t_2)^T; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^2 \sum_{j=1}^2 \pi_{ij} G_{SUN_{2,2}}((t_1, t_2)^T; \boldsymbol{\theta}_{ij})$, where $\pi_{11} = \pi_{22} = 0.2684$, $\pi_{12} = \pi_{21} = 0.2316$ and $\boldsymbol{\theta}_{ij} = (\boldsymbol{\xi}_{ij}, \boldsymbol{\eta}_{ij}, \boldsymbol{\Omega}_{ij}, \boldsymbol{\Gamma}_{ij}, \boldsymbol{\Lambda}_{ij})$. By some simple calculation we have for $i, j = 1, 2$

$$\boldsymbol{\xi}_{ij} = \begin{pmatrix} 1.65 \\ 3.5 \end{pmatrix}, \quad \boldsymbol{\eta}_{ij} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\Gamma}_{11} = \boldsymbol{\Gamma}_{22} = \begin{pmatrix} 0.04 & 0.004 \\ 0.004 & 0.03 \end{pmatrix},$$

$$\boldsymbol{\Gamma}_{12} = \boldsymbol{\Gamma}_{21} = \begin{pmatrix} 0.04 & -0.004 \\ -0.004 & 0.03 \end{pmatrix}, \quad \boldsymbol{\Omega}_{11} = \begin{pmatrix} .044 & .080 \\ .080 & .297 \end{pmatrix}, \quad \boldsymbol{\Omega}_{12} = \begin{pmatrix} .044 & .085 \\ .085 & .343 \end{pmatrix},$$

$$\boldsymbol{\Omega}_{21} = \begin{pmatrix} .047 & .087 \\ .087 & .297 \end{pmatrix}, \quad \boldsymbol{\Omega}_{22} = \begin{pmatrix} .047 & .096 \\ .096 & .343 \end{pmatrix}, \quad \boldsymbol{\Lambda}_{11} = \begin{pmatrix} -.041 & 0 \\ -.003 & -.007 \end{pmatrix},$$

$$\mathbf{\Lambda}_{12} = \begin{pmatrix} -.041 & 0 \\ -.043 & .011 \end{pmatrix}, \mathbf{\Lambda}_{21} = \begin{pmatrix} .009 & 0 \\ .003 & -.007 \end{pmatrix}, \mathbf{\Lambda}_{22} = \begin{pmatrix} .009 & 0 \\ .043 & .011 \end{pmatrix}.$$

References

- Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions, *Scandinavian Journal of Statistics*, **33**, 561-574.
- Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics, *Annals of Statistics*, **2**, 1034-1039.
- David, H. A. (1973). Concomitants of order statistics, *Bulletin of International Statistical Institute*, **45**, 295-300.
- Jamalizadeh, A. and Balakrishnan, N. (2012). Concomitants of order statistics from multivariate elliptical distributions, *Journal of Statistical Planning and Inference*, **142**, 397-409.
- Wang, Ke. and Nagaraja, H. N. (2009). Concomitants of order statistics for dependent samples, *Statistics and Probability Letters*, **79**, 553-558.



A study for testing Zero inflated Negative Binomial versus Negative Binomial distribution

Zahra Saberi¹, Maryam Hassanpour Khademi²

¹Assistant professor of statistics, Faculty of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran.

² M.S. of statistics, Faculty of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran.

Abstract: Count data often show a higher incidence of zero counts than would be expected if the data were negative binomial distributed. Zero-inflated negative binomial models are a useful class of models for such data, but parameter estimates may be seriously biased if the nonzero counts are over dispersed in relation to the negative binomial distribution. These types of data are commonly found in various scientific disciplines such as finance, insurance, biomedical, econometrical, ecology, and health sciences. We therefore provide a score test for testing negative binomial models against zero-inflated negative binomial alternatives. The score test has an advantage over the LRT and the Wald test in that the score test only requires that the parameter of interest be estimated under the null hypothesis. The purpose of this paper is test a negative Binomial against a zero inflated negative binomial model, we show that zero inflated negative binomial better fitting for count data with excess zero.

Keywords Count data, Negative binomial distribution, Score test, Zero inflation distribution, Zero inflated Negative Binomial distribution

Mathematics Subject Classification (2010): 62 – 07 62F15 62J12.

1 Introduction

Zero-inflated count models provide a method to explain the excess zeros by modeling the data as a mixture of two separate distributions: one distribution is typically a Poisson or negative binomial distribution that can generate both zero and nonzero counts, and the second distribution is a constant distribution that generates only zero counts. Standard discrete distribution may fail to fit such data either because of zero inflation or over/underdispersion. Now, there is increased interest in a zero inflated distribution to account for extra zeros in data (Xie et al. (2009)). Failure to account for extra zeros may result in biased parameter estimates and misleading inference. Because the zero inflated

²Maryam Hassanpour Khademi : maryam.hassanpour@math.iut.ac.ir

distributions usually provide better statistical fit, some researchers, e.g., [Hall \(2000\)](#), [Famoye and Singh \(2003\)](#), [Bodhisuwan \(2011\)](#) proposed these distributions.

As mentioned above, the mixed distribution defines one of the most important ways to obtain a new probability distribution in applied probability and operational research ([Gomez-Deniz et al. \(2008\)](#)). For the purpose, we are looking for a new zero inflated distribution which is a more flexible alternative to fit count data with excess zeros.

When the underlying count distribution is a negative binomial, the mixture is called a zero-inflated negative binomial (ZINB) distribution. ZINB models are better models to fit count data with excess zeros. For testing appropriate model to count data can use the test such as Wald test, likelihood ratio test and score test. The score test has an advantage over the LRT and the Wald test in that the score test only requires that the parameter of interest be estimated under the null hypothesis.

The ZINB model was discussed in [Ridout et al. \(2001\)](#), where a score test is provided for testing ZIP regression models against ZINB alternatives. A study of the score test in discrimination poisson and zero inflated poisson models provided by [Silva et al. \(2013\)](#). For independent count data with many zeros, [Van den Broek \(1995\)](#) have studied a score test for zero inflation in a Poisson distribution. [Deng and Paul \(2005\)](#) have presented score tests for zero inflation and over-dispersion in generalized linear models. [Xiang et al. \(2006\)](#) have developed a score test statistic for testing over-dispersion in the ZIP mixed regression model against the ZINB mixed regression model based on the best linear unbiased prediction type (BLUP) log likelihood.

In this paper, a score test provide for testing zero inflated negative Binomial against a negative binomial model and show that zero inflated negative binomial models fitted better for count data with excess zeros.

In Sections 2 gives an introduction to zero inflated negative binomial distribution and we said the score test for testing NB against ZINB are described. In section 3 a simulation study have done and the power and type I error of score test computed for different sample sizes and different values of parameters. At the end, the conclusion and discussion mentioned.

2 Score test for Zero Inflated Negative Binomial distribution

let Y be a discrete non-negative random variable with a zero inflated distribution, where ω denote the proportion of structural zeros:

$$P(Y = y) = \omega f_1(y) + (1 - \omega) f_2(y), \quad y = 0, 1, 2, \dots$$

where $0 < \omega < 1$, $f_2(y)$ is a negative binomial probability mass function (pmf) of Y with parameter r , p , and $f_1(y)$ denotes as

$$f_1(y) = \begin{cases} 1 & y = 0 \\ 0 & o.w. \end{cases}$$

Now, we introduce some zero inflated distributions are as follows. If its probability mass function is given by

$$f(Y_i, \omega, p) = \begin{cases} \omega + (1 - \omega)p^r & ; y_i = 0 \\ (1 - \omega) \binom{r+y_i-1}{y_i} p^r (1-p)^{y_i} & ; y_i \neq 0 \end{cases} \quad (2.1)$$

A random variable Y_i ($i=1, 2, \dots, n$) is said to be distribution according to a Zero Inflated Negative Binomial distribution with parameters r , p and ω , which we denote by $ZINB(r, p, \omega)$, where ω refers to the proportion of zeros, r number of successful and p is the success probability of the bernoulli distribution. The mean and variance of this distribution are $(1-\omega)r\frac{1-p}{p}$ and $(1-\omega)r\frac{1-p}{p^2}$, respectively.

Score tests are widely used for testing misspecifications in count models because they require to use the model only under the null hypothesis. Now consider score test for ZI parameter, ω , that is interpreted as the probability of obtaining a zero-out come from the Bernoulli distribution.

The test considers $\omega = 0$ against $\omega > 0$, where ω is the proportion of zeros contained in the population and it is the zero-inflation (deflation) parameter which can take negative values ($-\frac{f(0, \omega, p)}{f(0, \omega, p) - 1} < \omega < 0$). Note that a zero-inflated model has $\omega > 0$. On the null hypothesis ($\omega = 0$), (2.1) is reduced to the negative binomial distribution. The rejection of the null hypothesis of this test leads us to interpret the model to be used should be the Zero Inflated Negative Binomial (ZINB) distribution.

For doing this test, the GLR and Wald statistics cannot be used, since the MLE of parameters of this distribution hadn't closed form. So a score test used as a test statistic. Score test need to the first and second order partial differentiation and parameter(s) estimate under null hypothesis. So, let y_1, y_2, \dots, y_n be a random sample from (2.1) and the number of observations equal to zero was defined by n_0 . For the remaining sample values $y = 1, 2, \dots$, we used the notations n_1 to number of observations equal to 1, n_2 for the number of observations equal to 2 and so on, the $n = n_0 + \sum_{y=1}^{\infty} n_y$. Thus, the likelihood function is defined as,

$$\begin{aligned} L(\omega, p | y_i) &= \prod_{i=1}^n [(\omega + (1 - \omega)p^r)I_{y_i=0} + I_{y_i>0}(1 - \omega) \binom{r + y_i - 1}{y_i} p^r (1 - p)^{y_i}] \\ &= (\omega + (1 - \omega)p^r)^{n_0} \left(\prod_{i=1, y_i > 0}^n \binom{r + y_i - 1}{y_i} \right) (1 - \omega)^{n - n_0} p^{(n - n_0)r} (1 - p)^{\sum_{i=1, y_i > 0}^n y_i} \end{aligned} \quad (2.2)$$

and the log likelihood function can be written as:

$$l(\omega, p|y_i) = n_0 \log(\omega + (1 - \omega)p^r) + (n - n_0) \log(1 - \omega) + (n - n_0)r \log p \\ + \sum_{i=1, y_i > 0}^n y_i \log(1 - p) + \sum_{i=1, y_i > 0}^n \log \binom{r + y_i - 1}{y_i}.$$

So the first order partial derivatives were calculated using the log-likelihood function are,

$$\frac{dl(\omega, p|y_i)}{dp} = \frac{n_0 r (1 - \omega) p^{r-1}}{\omega + (1 - \omega) p^r} + \frac{(n - n_0)r}{p} - \frac{\sum_{i=1}^n y_i}{1 - p}, \quad (2.3)$$

$$\frac{dl(\omega, p|y_i)}{d\omega} = \frac{n_0(1 - p^r)}{\omega + (1 - \omega)p^r} - \frac{(n - n_0)}{1 - \omega}. \quad (2.4)$$

Where haven't any closed form for \hat{p} and $\hat{\omega}$, but under $H_0 : \omega = 0$, $\hat{p} = \frac{nr}{nr + \sum_{i=1}^n y_i}$. Then the score test for testing $H_0 : \omega = 0$ against $H_1 : \omega > 0$ is closed as;

$$S(0) = \frac{(dl(\omega, p | y_i)/d\omega)^2}{-E(\partial^2 l(\omega, p | y_i)/\partial \omega^2)} = \frac{(n_0 - np_0^r)^2}{n_0(1 - p_0^r)^2 + (n - n_0)p_0^{2r}}, \quad (2.5)$$

where $p_0 = \hat{p}$ is the MLE estimate under H_0 , n is the sample size and n_0 is the number of zero observations contained in the sample. This statistic defined under $H_0 : \omega = 0$ is given by Equation (2.5) that asymptotically converges to the chi-square distribution with one degree of freedom. Rejection of H_0 implies that ZINB model should be used to adjust the data distribution.

3 Monte Carlo Simulation

We examine the performances of score test statistics via Monte Carlo simulations to provide finite sample properties of the proposed statistics. Using the Monte Carlo simulation, it was carried out an assessment of the properties of the control of type I error and power of the score test so that the likelihood of type I error was determined by the proportion of times where the statistic S were rejected at nominal level of significance set at 5% when the sample were simulated on H_0 . Similarly the power was computed, considering the simulated samples of H_1 . Thus, the parametric values used in 10000 Monte Carlo iterations were specified as the following description: a) Negative Binomial distribution success probability (p) fixed at 0.6, 0.65, 0.7 and 0.75; b) different parametric values set by $\omega = 0.009, 0.01, 0.02, 0.03$ and 0.05 ; c) different sample sizes (n) set at $n = 5, 15, 50, 100$ and 150 , also for these simulations, are considered $r = 3$. Thus, the effective control of type I error provided by the test, considering a sample size that expresses high value of power, which is allowed to recommend the required sample size for the score test can be used in the discrimination of Negative binomial or ZINB models.

Table 1: Type I error for the score test as function of sample size (n), $r = 3$, successes probability of NB, nominal level of significance $\alpha = 5\%$ and $N = 10000$.

n	p			
	0.60	0.65	0.70	0.75
5	0.1186	0.0792	0.0580	0.0745
15	0.0709	0.0329	0.0380	0.0729
50	0.0397	0.0467	0.0464	0.0617
100	0.0636	0.0421	0.0555	0.0509
150	0.0534	0.0470	0.0467	0.0469

4 Results and discussion

The results of simulation studies are presented in tables 1,2. The results of that Table 1, indicated that, for all values of p , When the sample size increases, the probability of type I error was close to the nominal level of significance. Also, for small sample sizes, such as $n = 5, 15$, the score test is conservative. Score test power keeping the same parametric specifications regarding the negative binomial probability successful of (p) and different sample sizes, the score test power was evaluated under the alternative hypothesis considering the proportions of zero values hypothesized by at $\omega = 0.009, 0.01, 0.02, 0.03$ and 0.05 . The values obtained for the study of the power test are described in Table 2. However, it should be emphasized that according to the results on the probability of occurrence of Type I error control have been consistent with the nominal level of significance set at 0.05 . The results of Table 2 showed that the power increases slowly when ω increases, e.g. for successful probability $p = 0.75$ and $n = 50$, the power increases of 0.8327 to 0.9008 . The power of test increased closed to 1 in high successful probability and large sample size.

References

- Bodhisuwan, W. (2011). Zero inflated Waring distribution and its application. Proceedings of the 37th Congress on Science and Technology of Thailand, Thailand.
- Deng, D., Paul, S.R., (2005). Score tests for zero-inflation and over-dispersion in generalized linear models. *Statistica Sinica*, **15**, 257–276.
- Famoye, F. and Singh, K.P. (2003). On infated generalized Poisson regression models. *Advances and Applications in Statistics*, **3**, 145–158.

Table 2: power of the test score for nominal level of significance $\alpha = 5\%$ and depending on the percentage of zeros (ω), sample size (n), successful probability (p) and $r = 3$

(n, p)	ω				
	0.009	0.01	0.02	0.03	0.05
(5,0.6)	0.1618	0.1666	0.1621	0.1703	0.1684
(15,0.6)	0.5928	0.5958	0.6072	0.6337	0.6623
(50,0.6)	0.9941	0.9942	0.996	0.9967	0.9983
(100,0.6)	1	1	1	1	1
(150,0.6)	1	1	1	1	1
(5,0.65)	0.1688	0.1707	0.1738	0.1809	0.1809
(15,0.65)	0.4865	0.4888	0.5015	0.5667	0.5667
(50,0.65)	0.9844	0.9861	0.9871	0.9908	0.9935
(100,0.65)	0.999	0.9997	0.9998	0.9998	1
(150,0.65)	1	1	1	1	1
(5,0.7)	0.1797	0.1854	0.1908	0.1908	0.1953
(15,0.7)	0.3833	0.3911	0.4302	0.4302	0.4602
(50,0.7)	0.9347	0.9418	0.9435	0.9543	0.9657
(100,0.7)	0.9993	0.9988	0.993	0.9992	0.998
(150,0.7)	0.9998	1	1	1	1
(5,0.75)	0.1102	0.114	0.1185	0.1153	0.1259
(15,0.75)	0.3186	0.3141	0.3287	0.3418	0.3744
(50,0.75)	0.8327	0.8307	0.8457	0.8702	0.9008
(100,0.75)	0.9864	0.9878	0.9906	0.9947	0.9969
(150,0.75)	0.9991	0.9996	0.9997	1	0.9998

Famoye, F., Singh, K. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, **4**(1), 117-130.

Gomez-Deniz E., Sarabia, J.M. and Caldern-Ojeda, E. (2008). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications, *Insurance Mathematics and Economics*, **42**, 39-49.

Hall, D.B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, **56**, 1030-1039.

Ridout, M., Hinde, J., Demtrio, C.G.B., (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, **57**, 219-223.

Silva, A. M., Cirillo, M. A. (2010). Estudo por simulao monte carlo de um estimador robust utilizado na inferncia de um modelo binomial contaminado. *Acta Scientiarum. Technology*, **32**(3), 303-307.

- Vanessa Siqueira Peres da Silva, Marcelo Angelo Cirillo and Juliana Garcia Cespedes, (2013), Acta scientiarum, *Technology*, **35(2)**, 333–337.
- Van den Broek, J., (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, **51**, 738–743.
- Xiang, L., Lee, A.H., Yau, K.K.W., Mclachlan, G.J., (2006). A score test for zero- inflation in correlated count data. *Statistics in Medicine*, **25**, 1660–1671.
- Xie, F., Wei, B., (2009). Influence analysis for count data based on generalized Poisson regression models. *Statistics*, **99999 (1)**, 1–20.



The residual lifetime of linear consecutive k -out-of- n systems with exchangeable components

E. Salehi¹

¹Birjand University of Technology, Birjand, IRAN

Abstract: In this paper, we consider the linear consecutive k -out-of- n systems consisting of exchangeable components. We study the reliability properties of the residual lifetimes of such systems under the condition that at least $n - r + 1$, ($r \leq n$), components of the system are operating and present some stochastic comparisons on this conditional random variable.

Keywords Order statistics, Exchangeable, Stochastic ordering, Reliability.

Mathematics Subject Classification (2010): 62N05 60E15.

1 Introduction

The consecutive k -out-of- n and related systems are important structures of coherent systems in field of reliability engineering. The consecutive k -out-of- n systems consist of n components arranged in a line (linear case), and they fail (function) if a particular number of consecutive components fail (work). These systems are used for modelling various engineering systems such as microwave stations of a telecom network, oil pipeline systems, and vacuum systems in an electron accelerator. One of the most popular type of consecutive systems is consecutive k -out-of- n system. A linear consecutive k -out-of- n :F (G) (Lin/Con/ k/n :F (G)) system consists of n linearly arranged components such that the system fails (works) if and only if at least k consecutive components fail (work). For a review of consecutive type systems, one can see [Chang et al. \(2000\)](#) or [Kuo and Zuo \(2003\)](#).

The study on the reliability properties of consecutive k -out-of- n systems has been considered by various researchers. [Chen and Hwang \(1985\)](#) investigated the failure distribution of consecutive k -out-of- n :F systems in the case where the system has independent and identically distributed components. [Aki and Hirano \(1996\)](#) represented the reliability function of a linear consecutive k -out-of- n :F system as mixture of the reliability of ordered lifetime of its components, whenever the lifetimes are independent and identically distributed. Stochastic ordering results on the lifetime of linear and circular consecutive k -out-of- n :F systems have been obtained by [Boland and Samaniego \(2004\)](#). [Navarro and Eryilmaz \(2007\)](#) studied the mean residual lifetimes of the systems having exchangeable components. [Eryilmaz](#)

¹Speaker's Name: E. Salehi, salehi@birjandut.ac.ir

(2009) studied the reliability properties of linear consecutive k -out-of- n systems including arbitrarily dependent components. Recently, Salehi et al. (2011) and Salehi et al. (2012) investigated the residual lifetime of of consecutive k -out-of- n systems under different scenarios.

Suppose that T is the lifetime of a coherent system which includes n components with lifetimes of T_1, T_2, \dots, T_n . Denote by $T_{1:n}, T_{2:n}, \dots, T_{n:n}$ the ordered lifetimes of the components. The conditional random variable $\{T - t | T_{r:n} > t\}$ represents the residual lifetime of the system under the condition that at least $n - r + 1$ components of the system are working at time t . The residual lifetime of a system have been considered by many authors recently. Among others, we refer to Asadi and Bairamov (2006), Bairamov and Arnold (2007), Li and Zhang (2008), Li and Zhao (2008), Zhao et al. (2008) and Kochar and Xu (2010). Most of the presented results obtained in the literatures are restricted to the case when the components are independent and identically distributed. However, in fact the components are not independent in real, and then the researchers have been paid their attentions on this subject. In this regards, we can refer to Navarro et al. (2007), Navarro and Spizzichino (2010), Jia et al. (2010), Zhang (2010), Sadegh (2011), Navarro and Rubio (2011) and Rezapour et al. (2013).

This paper is organized as follows: In the beginning of Section 2, we give some preliminary results on distributional properties of Lin/Con/ k/n :G (F). In Section 3, we present two important lemmas to obtain the main results. In following, by using lemmas we obtain the reliability of the residual lifetime of the linear consecutive k -out-of- n including exchangeable components, and also, we do some comparisons between them in different stochastic orders.

2 Preliminaries

A system consist of n components is said to be a Lin/Con/ k/n :G(F) system such that the system operates (fails) if and only if at least k consecutive components operate (fail). Let $T_i, i = 1, 2, \dots, n$, denote the lifetimes of n components which are connected to the consecutive systems Lin/Con/ k/n :G or Lin/Con/ k/n :F. Denoting by $T_{k|n}^L$ the lifetime of the Lin/Con/ k/n system, it can be shown that the lifetimes of these systems can be represented as follows, respectively.

$$\begin{aligned} T_{k|n:G}^L &= \max\{T_{[1:k]}, T_{[2:k+1]}, \dots, T_{[n-k+1:n]}\}, \\ T_{k|n:F}^L &= \min\{T^{[1:k]}, T^{[2:k+1]}, \dots, T^{[n-k+1:n]}\}, \end{aligned}$$

where $T_{[i:m]} = \min\{T_i, \dots, T_m\}$ and $T^{[i:m]} = \max\{T_i, \dots, T_m\}$ for $1 \leq i < m \leq n$.

Eryilmaz (2009), for the case of $2k \geq n$, obtained the explicit form of the reliability function of the systems Lin/Con/ k/n :G and Lin/Con/ k/n :F as

$$\bar{F}_{k|n:G}^L(t) = P(T_{k|n:G}^L > t) = \sum_{i=k}^n [P(T_{[i-k+1:i]} > t) - P(T_{[i-k+1:i+1]} > t)], \quad (2.1)$$

and

$$\bar{F}_{k|n:F}^L(t) = P(T_{k|n:F}^L > t) = 1 - \sum_{i=k}^n [P(T_{[i-k+1:i]} \leq t) - P(T_{[i-k+1:i+1]} \leq t)], \quad (2.2)$$

where $P(T_{[n-k+1:n+1]} > t) = P(T_{[n-k+1:n+1]} \leq t) = 0$. In these representations one does not need to have the assumptions that the components lifetimes be either independent or identically distributed, i.e. the representations are valid if the components are dependent and have a arbitrary joint distribution.

In the following, we present the concept of usual stochastic order and important lemma, for more details we refer the reader to Shaked and Shanthikumar (2007).

Definition 2.1. Let X and Y be random variables with survival functions \bar{F} and \bar{G} , respectively. X is said to be smaller than Y in the usual stochastic order (denoted by $X \leq_{st} Y$) if $\bar{F}(t) \leq \bar{G}(t)$ for all t .

Lemma 2.2. Let $\{X_1, X_2, \dots\}$ be a sequence of (not necessarily independent) random variables. Then $X_{i:m} \leq_{st} X_{j:n}$, whenever $i \leq j$ and $m - i \geq n - j$.

3 Main results

Let $\mathbf{T} = (T_1, T_2, \dots, T_n)$ be a exchangeable random vector and assume that \mathbf{T} has an arbitrarily joint distribution function $F(t_1, t_2, \dots, t_n)$ with corresponding joint reliability function $\bar{F}(t_1, t_2, \dots, t_n)$. In the following we give two important lemmas which are required to present the main results (for proof see Eryilmaz (2011)).

Lemma 3.1. Let T_1, T_2, \dots, T_n be exchangeable lifetimes of n components with joint distribution $F(t_1, t_2, \dots, t_n)$. Then, for $x, t > 0$, the joint survival function of $T_{1:k}$ and $T_{r:n}$ is

$$\begin{aligned} H_{k,r,n}(x, t) &= P(T_{1:k} > x + t, T_{r:n} > t) \\ &= \sum_{j=\max\{k, n-r+1\}}^n \sum_{m=0}^{j-k} \binom{n-k}{j-k} \binom{j-k}{m} p_{m,j}(x, t), \end{aligned} \quad (3.1)$$

where

$$p_{m,j}(x,t) = \sum_{l_1=0}^m \sum_{l_2=0}^{n-j} (-1)^{l_1+l_2} \binom{m}{l_1} \binom{n-j}{l_2} \bar{F}(\underbrace{0, \dots, 0}_{n-j-l_2}, \underbrace{t, \dots, t}_{m-l_1+l_2}, \underbrace{x+t, \dots, x+t}_{j-m+l_1}).$$

Lemma 3.2. Let T_1, T_2, \dots, T_n be exchangeable lifetimes of n components with joint distribution $F(t_1, t_2, \dots, t_n)$. Then, for $x, t > 0$, the joint distribution function of $T_{k:k}$ and $T_{r:n}$ is

$$\begin{aligned} I_{k,r,n}(x,t) &= P(T_{k:k} \leq x+t, T_{r:n} \leq t) \\ &= \sum_{j=r}^n \sum_{m=\max\{0, k-j\}}^{\min\{k, n-j\}} \binom{k}{m} \binom{n-k}{j-k+m} q_{m,j}(x,t), \end{aligned} \tag{3.2}$$

where

$$q_{m,j}(x,t) = \sum_{l_1=0}^m \sum_{l_2=0}^j (-1)^{l_1+l_2} \binom{m}{l_1} \binom{j}{l_2} \bar{F}(\underbrace{0, \dots, 0}_{j-l_2}, \underbrace{t, \dots, t}_{m-l_1+l_2}, \underbrace{x+t, \dots, x+t}_{n-j-m+l_1}).$$

Now, let T_1, T_2, \dots, T_n be the exchangeable random variables representing the lifetimes of the components of the consecutive k -out-of- n system. Assume that $\mathbf{T} = (T_1, T_2, \dots, T_n)$ have the joint arbitrarily survival $\bar{F}(t_1, t_2, \dots, t_n)$. In the following, we represent the reliability function of the residual lifetime of linear consecutive k -out-of- n systems, i.e. $T_{n,G,L}^{k,r}(t) = \{T_{k|n:G}^L - t | T_{r:n} > t\}$ and $T_{n,F,L}^{k,r}(t) = \{T_{k|n:F}^L - t | T_{r:n} > t\}$.

Theorem 3.3. Let T_1, T_2, \dots, T_n be the exchangeable lifetimes of n components of Lin/Con/ k/n system. Then for $2k \geq n$ and $x, t > 0$,

$$P(T_{k|n:G}^L - t > x | T_{r:n} > t) = \frac{1}{\bar{F}_{r:n}(t)} \left[(n-k+1)H_{k,r,n}(x,t) - (n-k)H_{k+1,r,n}(x,t) \right], \tag{3.3}$$

and

$$\begin{aligned} P(T_{k|n:F}^L - t > x | T_{r:n} > t) &= 1 - \frac{1}{\bar{F}_{r:n}(t)} \left[(n-k+1) \left(1 - \bar{F}_{k:k}(x+t) - I_{k,r,n}(x,t) \right) \right. \\ &\quad \left. - (n-k) \left(1 - \bar{F}_{k+1:k+1}(x+t) - I_{k+1,r,n}(x,t) \right) \right], \end{aligned} \tag{3.4}$$

where $H_{k,r,n}(x,t)$ and $I_{k,r,n}(x,t)$ are defined in (3.1) and (3.2), respectively.

Proof. For $2k \geq n$,

$$P(T_{k|n:G}^L > x+t | T_{r:n} > t) = \frac{1}{\bar{F}_{r:n}(t)} P(T_{k|n:G}^L > x+t, T_{r:n} > t).$$

From (2.1), the exchangeable assumption, and using Lemma 3.1, we reach to expression (3.3). Similarly for Lin/Con/ $k/n:F$ system, from (2.2) and using Lemma 3.2, the proof is complete. \square

Theorem 3.4. Let T_1, T_2, \dots, T_n be the arbitrarily random variables showing the lifetimes of the components of a Lin/Con/ k/n : G system with a joint survival function $\bar{F}(t_1, t_2, \dots, t_n)$. Then for $k \geq \max\{n - r + 1, \lfloor \frac{n}{2} \rfloor\}$ and for all $t > 0$,

$$T_{n,G,L}^{r+1,k}(t) \leq_{st} T_{n,G,L}^{r,k}(t).$$

Proof. To prove the required result, we have to show that for $t > 0$,

$$P(T_{k|n:G}^L - t > x | T_{r+1:n} > t) \leq P(T_{k|n:G}^L - t > x | T_{r:n} > t), \text{ for all } x > 0.$$

Using (3.3), for $2k \geq n$, we have

$$P(T_{k|n:G}^L - t > x | T_{r:n} > t) - P(T_{k|n:G}^L - t > x | T_{r+1:n} > t) = \frac{\Delta(k) - \Delta(k + 1)}{\bar{F}_{r:n}(t)\bar{F}_{r+1:n}(t)},$$

where

$$\Delta(k) = (n - k + 1) \left[\bar{F}_{r+1:n}(t)H_{k,r,n}(x, t) - \bar{F}_{r:n}(t)H_{k,r+1,n}(x, t) \right], \quad 2k \geq n.$$

If $k \geq n - r + 1$, using (3.1), $H_{k,r,n}(x, t) = H_{k,r+1,n}(x, t)$ and $H_{k+1,r,n}(x, t) = H_{k+1,r+1,n}(x, t)$, for $l = 1, 2, \dots, n - k + 1$. Therefore,

$$\Delta(k) - \Delta(k + 1) = \left[\bar{F}_{r+1:n}(t) - \bar{F}_{r:n}(t) \right] \left[(n - k + 1)H_{k,r,n}(x, t) - (n - k)H_{k+1,r,n}(x, t) \right] \quad (3.5)$$

From Lemma 3.1, the first bracket in expression (3.5) is non-negative. Since $H_{k,r,n}(x, t) \geq H_{k+1,r,n}(x, t)$, when $k \geq n - r + 1$, the second bracket in expression (3.5) also is non-negative. Hence, for $k \geq n - r + 1$, $\Delta(k) - \Delta(k + 1) \geq 0$, and it means that $T_{n,G,L}^{r+1,k}(t) \leq_{st} T_{n,G,L}^{r,k}(t)$. Therefore, the proof is complete. \square

The following example gives an application of this theorem.

Example 3.5. Let T_1, T_2, T_3 and T_4 denote the exchangeable lifetimes of four components which are connected in a Lin/Con/3/4: G system and assume that T_i 's have a FGM multivariate exponential distribution with parameters $\theta = (\theta_{12}, \theta_{13}, \dots, \theta_{1234})$ and $\lambda > 0$, the joint survival function

$$\begin{aligned} &\bar{F}(t_1, t_2, t_3, t_4) \\ &= 1 + \sum_{p=1}^4 (-1)^p \sum_{1 \leq i_1 < i_2 < \dots < i_p \leq 4} \prod_{l=1}^p (1 - e^{-\lambda t_{i_l}}) \left[1 + \sum_{d=2}^p \sum_{1 \leq j_1 < j_2 < \dots < j_d \leq p} \theta_{j_1 j_2 \dots j_d} e^{-\lambda \sum_{l=1}^d t_{j_l}} \right], \end{aligned}$$

where $j_1, j_2, \dots, j_d \in \{i_1, i_2, \dots, i_p\}$, $i_1, i_2, \dots, i_p \in \{1, 2, 3, 4\}$. In special case, let $\theta_{12} = \theta_{13} = \dots = \theta_{1234} = \theta$, hence

$$\bar{F}(t, t, t, t) = 1 - 4F(t) + 6F^2(t) \left[1 + \theta \bar{F}^2(t) \right] - 4F^3(t) \left[1 + 3\theta \bar{F}^2(t) + \theta \bar{F}^3(t) \right]$$

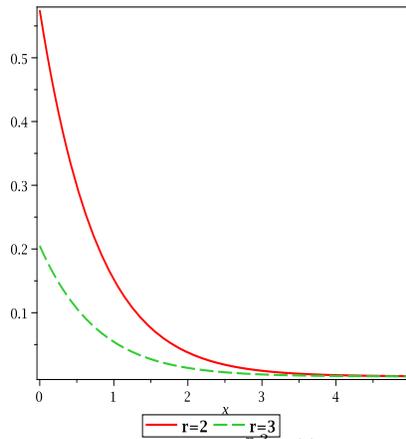


Figure 1: The curves of survival functions of $T_{4,G,L}^{r,3}(t)$, $r = 2, 3$ at $t = 1$, for Example 3.5.

$$+ F^4(t) \left[1 + 6\theta \bar{F}^2(t) + 4\theta \bar{F}^3(t) + \theta \bar{F}^4(t) \right].$$

It can be shown that

$$\begin{aligned} & P(T_{3|4:G}^L - t > x | T_{2:4} > t) - P(T_{3|4:G}^L - t > x | T_{3:4} > t) \\ &= \frac{\bar{F}_{3:4}(t) - \bar{F}_{2:4}(t)}{\bar{F}_{2:4}(t)\bar{F}_{3:4}(t)} \left[2\bar{F}(x+t, x+t, x+t) - \bar{F}(x+t, x+t, x+t, x+t) \right] \\ &\geq 0. \end{aligned}$$

The graphs of the survival functions of $T_{4,G,L}^{r,3}(t)$, $r = 2, 3$, for $\theta = 0.25$, $\lambda = 0.5$, at a fixed point $t = 2$, are plotted in Figure 1.

Remark 3.6. It is shown that when the linear consecutive k -out-of- n systems had arbitrary components then for all $t > 0$ (see Eryılmaz (2010))

$$T_{n,G,L}^{r,k+1}(t) \leq_{st} T_{n,G,L}^{r,k}(t), \quad T_{n,F,L}^{r,k}(t) \leq_{st} T_{n,F,L}^{r,k+1}(t).$$

References

- Aki, S. and Hirano, K. (1996). Lifetime distributions and estimation problems of consecutive k -out-of- n :F systems, *Ann. Inst. Stat. Math.*, **48**(1), 185-199.
- Asadi, M., and Bairamov, I. (2006). The mean residual life function of a k -out-of- n structure at the system level, *IEEE Trans. Reliab.*, **55**, 314-318.
- Bairamov, I. and Arnold, C. B., (2007). On the residual lifelengths of the remaining components in an $n - k + 1$ out of n system, *Stat. Prob. Lett.*, **78**, 945-952.
- Boland, P. J. and Samaniego, F. J. (2004). Stochastic ordering results for consecutive k -out-of- n :F systems, *IEEE Trans. Reliab.*, **53**(1), 7-10.
- Chang, G. J. , Cui, L. and Hwang, F. K. (2000). *Reliabilities of consecutive- k systems*, Kluwer Academic, Dordrecht.
- Chen, R. W. and Hwang, F. K. (1985). Failure distributions of consecutive k -out-of- n :F systems, *IEEE Trans. Reliab.*, **34**, 338-341.
- Eryilmaz, S. (2009). Reliability properties of consecutive k -out-of- n systems of arbitrarily dependent components, *Reliab. Eng. Syst. Saf.*, **94**, 50-356.
- Eryilmaz, S. (2010). Conditional lifetimes of consecutive k -out-of- n systems, *IEEE Trans. Reliab.*, **59**, 178-182.
- Eryilmaz, S. (2011). Circular consecutive k -out-of- n systems with exchangeable dependent components, *J. Statist. Plann. Infer.*, **141**, 725-733.
- Jia, X., Cui, L. and Yan, J. (2010). A study on the reliability of consecutive k -out-of- n : G systems based on copula, *Comm. Stat. Theory Meth.*, **39**, 2455-2472.
- Kochar, S. and Xu, M. (2010). On results lifetimes of k -out-of- n systems with nonidentical components, *Probab. Eng. Inform. Sci.*, **24**, 109-127.
- Kuo, W. and Zuo, M. J. (2003). *Optimal Reliability Modeling, Principles and Applications*, John Wiley and Sons, New York.
- Li, X. and Zhang, Z. (2008). Some stochastic comparisons of conditional coherent systems, *Appl. Stoch. Models Bus. Ind.*, **24**, 541-549.

- Li, X. and Zhao, P. (2008). Stochastic comparison on general inactivity time and general residual life of k -out-of- n : G systems, *Commun. Statist. Simul. Comput.*, **37**, 1005-1019.
- Navarro, J. and Eryilmaz, S. (2007). Mean residual lifetimes of consecutive k -out-of- n systems, *J. Appl. Probab.*, **44**, 82-98.
- Navarro, J. and Rubio, R. (2011). A note on necessary and sufficient conditions for ordering properties of coherent systems with exchangeable components, *Naval Res. Logist.*, **58**, 478489.
- Navarro, J., Ruiz, J. M. and Sandoval, C. J. (2007). Properties of coherent systems with dependent components, *Comm. Stat. Theory Meth.*, **36**, 1-17.
- Navarro, J. and Spizzichino, F. (2010). Comparisons of series and parallel systems with components sharing the same copula, *Appl. Stoch. Model. Bus. Indu.*, **26**, 775-791.
- Rezapour, M., Salehi, E. T., and Alamatsaz, M. H. (2013). Stochastic comparison of residual and past lifetimes of $(n - k + 1)$ -out-of- n systems with dependent components, *Comm. Stat. Theory Meth.*, **42**, 1-15.
- Sadegh, M. K. (2011). A note on the mean residual life function of a coherent system with exchangeable or nonidentical components, *J. Statist. Plann. Infer.*, **141**, 3267-3275.
- Salehi, E. T. , Asadi, M. and Eryilmaz, S. (2011). Reliability analysis of consecutive k -out-of- n systems with non-identical components lifetimes, *J. Statist. Plann. Infer.*, **141**, 2920-2932.
- Salehi, E. T. , Asadi, M. and Eryilmaz, S. (2012). On the mean residual lifetime of consecutive k -out-of- n systems, *Test*, **21**, 93-115.
- Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*, Springer, New York.
- Zhang, Z. (2010). Ordering conditional general coherent systems with exchangeable components, *J. Statist. Plann. Infer.*, **140**, 454-460.
- Zhao, P., Li, X., and Balakrishna, N. (2008). Conditional ordering of k -out-of- n systems with independent but non-identical components, *J. Appl. Probab.*, **45**, 1113-1125.



Expressions for the mean of the order statistics from the skew-normal distribution and their application

Mahdi Salehi ¹ and Mahdi Doostparast

Department of Statistics, Ferdowsi University of Mashhad,
P.O. Box 1159, Mashhad, 91775 Iran

Abstract: In design of experiments and reliability analyze, order statistics (OS) are used for various purposes including model checking, estimation of parameters and prediction. Most of these procedures are defined on the basis of the expectations of OS. In this paper, explicit expressions for moments of OS coming from the skew-normal (SN) distribution are derived. The SN model enjoys interesting properties from the normal distribution while captures asymmetric behaviour in the parent population. Two real data sets are analyzed as illustration purposes.

Keywords Order statistics, Skew-normal distribution, Goodness-of-fit test.

1 Introduction

Suppose that Y_1, \dots, Y_n are independent and identically distributed (iid) random variables. The corresponding order statistics (OS) are the Y_i 's arranged in increasing order of magnitude, denoted by $Y_{1:n} \leq \dots \leq Y_{n:n}$. The OS arise in many practical situations such as characterizations of probability distributions, goodness-of-fit tests, analyze of censored samples and reliability theory. For more details, see [David and Nagaraja \(2003\)](#) and the references therein.

Statistical inferences based on order statistics have been considered under some well-known parametric distributions such as exponential, Weibull, Burr, Pereto and normal distributions. Here in this paper, we are going to derive the moments of order statistics from the *skew-normal distribution*. A random variable Z_λ is said to have a standard skew-normal distribution with parameter $\lambda \in R$, denoted by $Z_\lambda \sim SN(\lambda)$, if its density is [[Azzalini \(1985\)](#)]

$$\phi_{SN}(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \text{Re}, \lambda \in \text{Re}, \quad (1.1)$$

where $\phi(z)$ and $\Phi(z)$ stand for the density and the cumulative distribution function (cdf) of the standard normal distribution, respectively.

¹Speaker: Salehi2sms@gmail.com (M. Salehi), doostparast@um.ac.ir (M. Doostparast)

The result obtained in this paper can be used in various statistical problems involving the normal distribution. We will illustrate the application of the obtained result in the goodness of fit tests. Therefore, the rest of this paper is organized as follows: Explicit expressions for moments of OS arising from the SN-distribution are derived in Section 2. In Section 3, two real data sets are analyzed. Section 4 concludes. The proofs are eliminated.

2 OS from the SN-distribution

In the sequel, let $Z_{i:n}$ be the i th order statistic in a random sample of size n coming from the standard normal distribution. The problems on moments of the $Z_{i:n}$ have been considered in the literatures. Among of them, Harter (1970) provided approximations for $E(Z_{i:n})$ ($n \leq 400$) and Bose and Gupta (1959) obtained explicit expressions for $E(Z_{i:n})$ when $n \leq 5$. Tietjen et al. (1977) also derived means, variances, and covariances of OS for sample sizes up to 50.

Here, we extend previous works and derive expression for moments of OS coming from the SN-distribution. To do this, let $Z_{(i:n),\lambda}$ be the i th order statistic from a random sample of size n coming from the $SN(\lambda)$ -distribution. Then, after using the binomial expansion, the density of $Z_{(i:n),\lambda}$ is

$$\phi_{i:n}(z; \lambda) = i \binom{n}{i} \phi_{SN}(z; \lambda) \sum_{d=0}^{n-i} \binom{n-i}{d} (-1)^d \Phi_{SN}^{d+i-1}(z; \lambda), \quad z \in \text{Re}, \tag{2.1}$$

for $i = 1, \dots, n$.

Chiogna (1998) proved that

$$E(Z_{(2:2),\lambda}) = \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}} + \left(\frac{2}{\pi}\right)^{1.5} \left(\sqrt{2} \tan^{-1} \sqrt{1+\lambda^2} - \frac{\lambda}{\sqrt{1+\lambda^2}} \tan^{-1} \frac{\lambda}{\sqrt{2}} \right). \tag{2.2}$$

Moreover, she derived a recurrence relation for the second moment of $Z_{(i:n),\lambda}$, $1 \leq i \leq n$. In this section, we first obtain an explicit expression for the moment generating function (*MGF*) of the the random variable $Z_{(i:n),\lambda}$, i.e. $M_{Z_{(i:n),\lambda}} = E(\exp\{tZ_{(i:n),\lambda}\})$, as a linear combination of the cdf of the multivariate normal distribution. First, we need the following lemma which is a special case of Lemma 3 of Jamalizadeh and Balakrishnan (2009). Hereafter, $\mathbf{0}_n$ is a column vector of all zeros of order n , $\mathbf{1}_n$ is a column vector of all ones of order n , and \mathbf{I}_n is the identity matrix of order n .

Lemma 2.1. *If $Z_\lambda \sim SN(\lambda)$ and $k(\geq 2)$ is a positive integer value, then*

$$\begin{aligned} E[Z_\lambda \Phi_{SN}^k(Z_\lambda; \lambda)] &= \frac{k2^{k+1}}{\pi^{1.5}} \cos^{-1} \left(\frac{-\lambda^2}{2+\lambda^2} \right) \Phi_{2k}(\mathbf{0}_{2k}; \tilde{\mathbf{\Omega}}_1) \\ &\quad + \frac{2^{k+1}\lambda}{\sqrt{2\pi(1+\lambda^2)}} \Phi_{2k}(\mathbf{0}_{2k}; \tilde{\mathbf{\Omega}}_2), \end{aligned} \tag{2.3}$$

where

$$\tilde{\Omega}_1 = \begin{pmatrix} \mathbf{I}_{k+1} & -\Delta_{k,\lambda}^{(1)T} \\ -\Delta_{k,\lambda}^{(1)} & \Sigma_{k-1,1} \end{pmatrix}, \quad \tilde{\Omega}_2 = \begin{pmatrix} \mathbf{I}_k & -\Delta_{k,\lambda}^{(2)T} \\ -\Delta_{k,\lambda}^{(2)} & \Sigma_{k,\lambda} \end{pmatrix}, \quad (2.4)$$

with partitioned covariance matrixes

$$\Delta_{k,\lambda}^{(1)} = \begin{pmatrix} \frac{\lambda}{\sqrt{1.5(1+\lambda^2)}} \mathbf{I}_{k-1} & \vdots & -\frac{\lambda}{\sqrt{6+3\lambda^2}} \mathbf{1}_{k-1} & \vdots & -\frac{\lambda}{\sqrt{6+3\lambda^2}} \mathbf{1}_{k-1} \end{pmatrix}_{(k-1) \times (k+1)}, \quad (2.5)$$

$$\Delta_{k,\lambda}^{(2)} = \frac{\lambda}{\sqrt{2+\lambda^2}} \mathbf{I}_k, \quad (2.6)$$

and

$$\Sigma_{k,\lambda} = \frac{1}{2+\lambda^2} (\mathbf{1}_k \mathbf{1}_k^T + [1+\lambda^2] \mathbf{I}_k). \quad (2.7)$$

In addition, for $k = 1$, the expression (2.3) simplifies to

$$E [Z_\lambda \Phi_{SN}(Z_\lambda; \lambda)] = \frac{2}{\pi^{1.5}} \tan^{-1} \sqrt{1+\lambda^2} + \frac{\lambda}{\sqrt{2\pi(1+\lambda^2)}} \left(1 - \frac{2}{\pi} \tan^{-1} \frac{\lambda}{\sqrt{2}} \right). \quad (2.8)$$

Remark 2.2. Upon substituting $\lambda = 0$ into (2.3)–(2.8) we get

$$E [Z \Phi^k(Z)] = \frac{k}{2\sqrt{\pi}} \Phi_{k-1}(\mathbf{0}_{k-1}; \Sigma_{k-1,1}), \quad k \geq 1, \quad (2.9)$$

where $Z \sim N(0, 1)$ and $\Phi_0(\cdot; \cdot) \equiv 1$.

Proposition 2.3. The MGF of the random variable $Z_{(i:n),\lambda}$ is

$$M_{Z_{(i:n),\lambda}}(s; \lambda) = 2^i i \binom{n}{i} e^{\frac{s^2}{2}} \sum_{d=0}^{n-i} (-2)^d \binom{n-i}{d} \Phi_{2j+1}(\tilde{\mathbf{y}}; \tilde{\Omega}), \quad (2.10)$$

where $j = d + i - 1$ and

$$\gamma = \begin{pmatrix} \mathbf{0}_j \\ \cdots \\ \frac{\lambda s}{\sqrt{1+\lambda^2}} \end{pmatrix}, \quad \Delta = \frac{\lambda}{\sqrt{2+2\lambda^2}} \begin{pmatrix} \mathbf{I}_j & \vdots & -\mathbf{1}_j \end{pmatrix}_{j \times (j+1)}, \quad (2.11)$$

$$\tilde{\mathbf{y}} = \begin{pmatrix} \gamma \\ \cdots \\ \frac{s}{\sqrt{2}} \mathbf{1}_j \end{pmatrix}, \quad \tilde{\Omega} = \begin{pmatrix} \mathbf{I}_{j+1} & -\Delta^T \\ -\Delta & \frac{1}{2} [\mathbf{1}_j \mathbf{1}_j^T + \mathbf{I}_j] \end{pmatrix}. \quad (2.12)$$

By differentiation from (2.10) at zero, we can derive $E\left(Z_{(i:n),\lambda}^k\right)$ for $k=1,2, \dots$. In this regard, Lemma 2 of Salehi et al. (2014) is useful. Notice that the mentioned method needs time-consuming algebraic manipulations. So, in the rest of this section we obtain the first moment of the random variable $Z_{(i:n),\lambda}$ directly from the density (2.1). To this end, we have

$$\begin{aligned} E[Z_{(i:n),\lambda}] &= i \binom{n}{i} \sum_{d=0}^{n-i} \binom{n-i}{d} (-1)^d \int_{-\infty}^{+\infty} z \phi_{SN}(z; \lambda) \Phi_{SN}^{d+i-1}(z; \lambda) dz \\ &= i \binom{n}{i} \sum_{d=0}^{n-i} \binom{n-i}{d} (-1)^d E[Z_\lambda \Phi_{SN}^{d+i-1}(Z_\lambda; \lambda)], \quad i = 1, \dots, n. \end{aligned} \tag{2.13}$$

So, using the fact that $E(Z_\lambda) = \lambda \sqrt{2/(\pi(1+\lambda^2))}$ and Lemma 2.1, the explicit expressions for $E[Z_{(i:n),\lambda}] (i = 1, \dots, n)$ are obtained.

Example 2.4. For $n = i = 2$ in (2.13) we have

$$\begin{aligned} E(Z_{(2:2),\lambda}) = 2E[Z_\lambda \Phi_{SN}(Z_\lambda; \lambda)] &= \frac{4}{\pi^{1.5}} \tan^{-1} \sqrt{1+\lambda^2} \\ &+ \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}} \left(1 - \frac{2}{\pi} \tan^{-1} \frac{\lambda}{\sqrt{2}}\right), \end{aligned}$$

which is equivalent to (2.2) derived by Chiogna (1998). Moreover, substituting $n = 2$ and $i = 1$ into (2.13) and using Lemma 2.1, we have

$$\begin{aligned} E(Z_{(1:2),\lambda}) &= 2 \left\{ E(Z_\lambda) - E[Z_\lambda \Phi_{SN}(Z_\lambda; \lambda)] \right\} \\ &= \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}} - \frac{4}{\pi^{1.5}} \tan^{-1} \sqrt{1+\lambda^2} - \left(\frac{2}{\pi}\right)^{1.5} \frac{\lambda}{\sqrt{1+\lambda^2}} \tan^{-1} \frac{\lambda}{\sqrt{2}}. \end{aligned}$$

Example 2.5. By substituting $\lambda = 0$ into (2.13) and then using (??), a close form for $E(Z_{i:n}) (i = 1, \dots, n)$ is obtained in term of the cdf of the normal distribution, i.e.

$$E[Z_{i:n}] = i \binom{n}{i} \sum_{d=0}^{n-i} \binom{n-i}{d} (-1)^d \frac{k}{2\sqrt{\pi}} \Phi_{k-1}(\mathbf{0}_{k-1}; \boldsymbol{\Sigma}_{k-1,1}), \tag{2.14}$$

where $k = d + i - 1$ and $\boldsymbol{\Sigma}_{k-1,1} = \frac{1}{3} \mathbf{1}_k \mathbf{1}_k^T + \frac{2}{3} \mathbf{I}_k$.

In Table 1, we present numerical values of $E[Z_{(i:15),\lambda}]$ given by (2.13) and (2.14) for some selected values of λ and i . We mentioned that the cdf of the normal distribution is computed using the statistical software R version 2.15.2 (see, R Development Core Team (2012)) with the package `mnormt`. Notice

that the $SN(\lambda)$ -distribution converges to the positive standard half-normal-distribution (HN^+) with density $\phi_{HN}^+(z) = 2\phi(x)$, $x \geq 0$, as λ goes to infinity. Therefore, the last row of Table 1 is associated with the expectation of the i th OS in a sample of size 15 coming from the HN^+ -distribution.

Corollary 2.6. *Since $-Z_\lambda \stackrel{d}{=} Z_{-\lambda}$, we have $E [Z_{(i:n),-\lambda}] = -E [Z_{(n-i+1:n),\lambda}]$.*

According to Corollary 2.6, the values of Table 1 can be used for the means of $Z_{(i:15),\lambda}$ for $\lambda \leq 0$.

Table 1: The values of $E [Z_{(i:15),\lambda}]$ given by Equation (2.13) for some selected values of λ .

λ	i							
	1	2	3	4	5	6	7	8
0	-1.7359	-1.2479	-0.9477	-0.7149	-0.5157	-0.3353	-0.1653	0.0000
1	-0.8273	-0.4499	-0.2145	-0.0303	0.1287	0.2739	0.4118	0.5470
2	-0.3771	-0.1093	0.0631	0.2018	0.3242	0.4388	0.5502	0.6620
3	-0.1925	0.0172	0.1565	0.2722	0.3774	0.4784	0.5790	0.6823
4	-0.1017	0.0740	0.1948	0.2984	0.3950	0.4898	0.5861	0.6865
5	-0.0503	0.1032	0.2135	0.3097	0.4017	0.4936	0.5882	0.6875
6	-0.0180	0.1191	0.2244	0.3148	0.4046	0.4951	0.5888	0.6878
7	0.0026	0.1316	0.2281	0.3183	0.4060	0.4957	0.5891	0.6879
∞	0.0787	0.1583	0.2393	0.3221	0.4075	0.4962	0.5893	0.6879

λ	i						
	9	10	11	12	13	14	15
0	0.1653	0.3353	0.5157	0.7149	0.9477	1.2479	1.7359
1	0.6832	0.8246	0.9760	1.1450	1.3450	1.6070	2.0428
2	0.7774	0.9000	1.0346	1.1884	1.3749	1.6251	2.0508
3	0.7911	0.9088	1.0398	1.1912	1.3762	1.6255	2.0509
4	0.7934	0.9100	1.0403	1.1914	1.3763	1.6256	2.0509
5	0.7939	0.9102	1.0404	1.1914	1.3763	1.6256	2.0509
6	0.7940	0.9102	1.0404	1.1914	1.3763	1.6256	2.0509
7	0.7940	0.9102	1.0404	1.1914	1.3763	1.6256	2.0509
∞	0.7941	0.9102	1.0404	1.1914	1.3763	1.6256	2.0509

3 Illustrative examples

To demonstrate the results obtained in preceding sections, we here analyzed two real data sets. In the examples, we used the procedure `sn.mle` in package `sn` of the statistical software R version 2.15.2 (R Development Core Team (2012)) to obtain the maximum likelihood (ML) estimates of the SN-distribution. For the Kolmogorov-Smirnov (KS) goodness-of-fit test, the procedure `ks.test` is implemented. For comparison various models, we considered the well known Akaike's Information Criterion (AIC) defined by $AIC = -2\log L + 2M$ where L and M , respectively, are the likelihood function (LF) and the number of the estimated parameters.

3.1 Times of electrical equipments data set

Here, we consider a data set on life testing of an given electrical equipment, due to [Nadarajah \(2008\)](#), planned for quality control purposes. The data are 15.17, 19.87, 20.18, 21.5, 21.88, 22.23, 23.02, 23.9, 28.17, 29.7. We assume that the lifetime random variable X follows the $SN(\lambda)$ -distribution with the location and the scale parameters μ and σ , respectively, i.e. $X \stackrel{d}{=} \mu + \sigma Z_\lambda$, where $Z_\lambda \sim SN$. The ML estimates of the parameters are obtained as $\hat{\mu}=19.2122$, $\hat{\sigma}=5.1563$ and $\hat{\lambda}=1.4002$ with $AIC=61.6165$. The p-value of the KS test is 0.9496 which support the adequacy of the SN -distribution. For an informal goodness of fit test, we plot the sorted observed failure times versus expectations of OS arising from $SN(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ obtained from (2.13), i.e. $\hat{\mu} + \hat{\sigma}E\left(Z_{(i:10),\hat{\lambda}}\right)$, for $i = 1, \dots, 10$, in Figure 1.

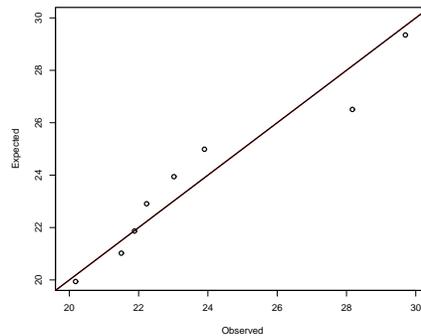


Figure 1: Plot of $\hat{\mu} + \hat{\sigma}E\left(Z_{(i:10),\hat{\lambda}}\right)$ ($i=1, \dots, 10$) versus the observed failure times of electrical equipments data set.

Notice that, [Nadarajah \(2008\)](#) fitted the gamma distribution by density $f(x; \alpha) = x^{\alpha-1}e^{-x}/\Gamma(\alpha)$ on the transformed data from x_i to $\bar{x}x_i/s_x^2$ with $AIC=63.3008$. So, by comparing AICs, the SN -distribution overcomes the gamma distribution for modeling this data set .

3.2 Fatigue data set

[Castillo et al. \(2005\)](#) considered a data set on occurrence of fatigue on a unit. They argued the suitability of the log-normal distribution. Assuming the log- SN -distribution for the lifetime of occurrence of fatigue T , i.e. $\log T \stackrel{d}{=} \mu + \sigma Z_\lambda$, the ML estimates of the parameters are derived as $\hat{\mu} = -2.7787$, $\hat{\sigma} = 0.5733$ and $\hat{\lambda} = 1.2896$. In this case, the p-value of the KS test and AIC are 0.9998 and 48.3492, respectively. We plotted the expected values $\hat{\mu} + \hat{\sigma}E\left(Z_{(i:35),\hat{\lambda}}\right)$ ($i=1, \dots, 35$) obtained from (2.13) versus the sorted observed data in Figure 2 which supports the suitability of the log- SN model for this data set.

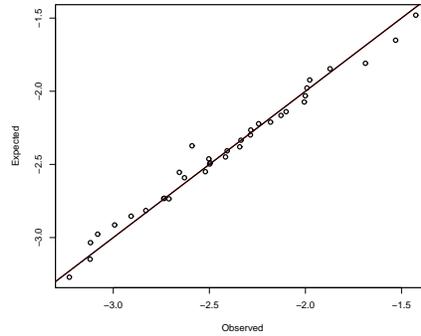


Figure 2: Plot of $\hat{\mu} + \hat{\sigma}E\left(Z_{(i:35),\lambda}\right)$ ($i=1, \dots, 35$) versus the sorted occurrence of fatigue

Conclusion

Order statistics coming from the SN -distribution were considered and explicit expressions for the mean and MGFs of OS were derived in terms of the cdf of a multivariate normal distribution. Our findings extended results of Chiogna (1998). To demonstrate applications of the obtained results, two real data sets were analysed.

References

- Azzalini, A. (1985). A class of distributions which includes the Normal ones, *Scand. J. Statist.*, **12**, 171–178.
- Bose, R. C. and Gupta, S. S. (1959). Moments of order statistics from a normal population, *Biometrika*, **46**, 433–440.
- Castillo, E., Hadi, S., Balakrishnan, N., Sarabia, JM. (2005). Extreme value and related models with applications in engineering and science. Wiley, Hoboken, NJ.
- Chiogna, M. (1998). Some results on the scalar skew-normal distribution, *Journal of the Italian Statistical Society*, **1**, 1–13.
- David, H. A. and Nagaraja, H. N. (2003). Order Statistics, Third edition, John Wiley & Sons, Hoboken, New Jersey.
- Harter, H. L. (1970). Order Statistics and Their Use in Testing and Estimation (Vol. 1 and 2), U.S. Government Printing Office, Washington, D.C.
- Jamalizadeh, A. and Balakrishnan, N. (2009). Order statistics from trivariate normal and t_ν -distributions in terms of generalized skew-normal and skew- t_ν distributions, *Journal of Statistical Planning and Inference*, **139**, 3799–3819.
- Nadarajah, S. (2008). Explicit expressions for moments of order statistics, *Statistics & Probability Letters*, **78**, 196–205.
- Salehi, M., Jamalizadeh, A., Doostparast, M. (2014). A generalized skew two-piece skew-elliptical distribution, *Statistical papers*, **55**, 409–429.
- R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Tietjen, G.L., Kahaner, D. K., and Beckman, R. J. (1977). Variances and covariances of the normal order statistics for sample sizes 2 to 50, *Selected Tables in Mathematical Statistics*, **5**, 1–73. 7, 122–128



Dirichlet processes applied to clustering of longitudinal data via piecewise mixed-effects models

Elaheh Shams¹, Reyhaneh Rikhtehgaran^{1,1}

¹Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran.

Abstract: This paper applies Dirichlet processes (DPs) in fitting piecewise mixed-effects models. The skew normal is adopted as the distribution of residuals for observations measured before a random time point and then the distribution shifts to the normal. The DP is considered as the prior for the distribution of the random change point. The discreteness nature of the DP is utilized to cluster subjects according to their distributional behavior over time along with achieving more efficient estimates. We also have used transition mixed-effects models to account for the autocorrelation among observations over time. A joint modeling approach is used to handle the bias created in these models. The Gibbs sampling technique is adopted to achieve the parameter estimates. The performance of the proposed model is illustrated by conducting a simulation study.

Keywords Dirichlet processes, Hierarchical models, Markov chain Monte Carlo simulation methods, Piecewise regression model, Skew-normal distribution.

Mathematics Subject Classification (2010): 62M99 62J05 62G07.

1 Introduction

Longitudinal studies are designed by repeatedly measuring subjects through time. The piecewise or switching regression models are often used in a wide variety of applications to study time trends in longitudinal data. These models can provide a reasonable statistical model if the underlying population which generates the data has structural changes at some time points. Situations where a shift takes place in the location parameter of the distribution have been studied before (e.g., [MacLain and Albert \(2014\)](#)).

However, in some applications, it usually happens that after certain times lag, shape of the distribution of response will change. This assumption corresponds to a switching regression model, in which the first part is related to a certain distribution for the errors and the second part is related to a different distribution. As a special case, we have assumed the normal and the skew-normal, respectively, as symmetric and asymmetric distributions for each part of the model. In practice, changes

¹Reyhaneh Rikhtehgaran : r_rikhtehgaran@cc.iut.ac.ir

usually happen at different time points for different subjects. Therefore, it is more realistic to assume a random change-point in the structure of mixed-effects models. In some applications, specifically in epidemiological studies, it is usually important that different subjects can be categorized based on behavior of their responses over time. In these studies, it usually takes time that a treatment including a new medicine, shows its symptoms and then before and after the time of change point the behavior of response is different. Addressing these issues, we propose the use of the Dirichlet process (DP) in the structure of transition mixed-effects models. The DP introduces an unknown distribution G over the space of all possible distribution functions (Ferguson (1973)). The DP has two parameters, a base distribution, G_0 , stating our guess about true non-parametric shape of G , and a precision parameter, M , reflecting our belief about how similar G is to G_0 . The DP generates discrete distributions, which can be useful for the purpose of data clustering.

To study the association of responses over time, we use transition mixed-effects models by assuming lagged responses in the mean structure of the model. We propose handling bias created in transition models by joint modeling of the initial responses and subsequent responses.

The remainder of this paper is organized as follows. In Sections 2 and 3, we briefly introduce the DP prior and the skew-normal distribution, respectively. Section 4 considers the specification of the proposed piecewise transition mixed-effects model. In Section 5, we implement the Gibbs sampling scheme. Section 6 presents a simulation study. The last section includes conclusions.

2 The Dirichlet process prior

This section introduces shortly the Dirichlet process. The DP introduced by Ferguson (1973) and developed by others, generates a discrete random probability measure G as $G(\cdot) = \sum_{j=1}^{\infty} \pi_j \Delta_{\xi_j}(\cdot)$, where $\Delta(\cdot)$ denotes a degenerated distribution of unit mass centered at ξ_j with $\xi_j \stackrel{\text{iid}}{\sim} G_0$ and $\pi_j = \gamma_j \prod_{i=1}^{j-1} (1 - \gamma_i)$ with $\gamma_j \stackrel{\text{iid}}{\sim} \text{Beta}(1, M)$ (we say $\pi_j \sim SB(M)$). This representation is called stick-breaking. We denote the distribution of G by $DP(M, G_0)$. Variation of G around G_0 is controlled by the dispersion parameter $M(> 0)$ in a stochastic way such that, if M gets large then G becomes close to G_0 . Being applicable, the above infinite summation is truncated to a finite integer C .

An attractive feature of the DP is to specify it as a prior for the distribution of random effects. Let $Y_i | \xi_i \stackrel{\text{iid}}{\sim} f(\cdot | \xi_i)$, $\xi_i | G \stackrel{\text{iid}}{\sim} G$ and $G \sim DP(M, G_0)$ for $i = 1, \dots, n$. The realizations of the DP are discrete, thus several ξ_i 's prone to take similar values, such that the number of distinct values of ξ_i , each constructing a cluster, being less than or equal to n . Each distinct value of ξ_i 's specifies a cluster. The number of clusters, K , depends on M , which controls the amount of clustering between the center effects.

2.1 The skew-normal distribution

The density function of the skew-normal distribution (Sahu et al. (2003)) with the location, scale and shape parameters μ , σ^2 and δ , respectively, denoted by $SN(y|\mu, \sigma^2; \delta)$, is given by

$$f(y) = \frac{2}{\sqrt{\sigma^2 + \delta^2}} \varphi\left(\frac{y - \mu}{\sqrt{\sigma^2 + \delta^2}}\right) \times \Phi\left(\frac{\delta}{\sigma^2} \frac{y - \mu}{\sqrt{\sigma^2 + \delta^2}}\right), \quad (2.1)$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the probability density and the cumulative distribution functions of $N(0, 1)$. For $\delta = 0$, the normal distribution is retrieved; for positive and negative values of δ , respectively, right and left skewed distributions are obtained. The skew normal distribution is shown to be in a hierarchical form $Y|z \sim N(\mu + \delta z, \sigma^2)$ where $z \sim N(0, 1)I\{z > 0\}$. This hierarchical representation is a powerful tool for the data analysis with computation being performed by the Markov chain Monte Carlo techniques.

3 The piecewise transition mixed-effects model

Let Y_{it} denotes the t -th measurement taken on the i -th individual, $t = 1, \dots, T$, $i = 1, \dots, n$. The proposed piecewise transition mixed-effects model for longitudinal data is given by

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}, \quad (3.1)$$

where vector $\boldsymbol{\beta}$ includes regression coefficients. In this model, the α_i 's denote random intercepts and the ε_{it} 's are residual terms. The covariate $y_{i,t-1}$ is the lagged-response variable for the i -th subject which represents state dependence. It is usually assumed that random effects and ε_{it} 's are independent for all i and t , and $\alpha_i \stackrel{\text{iid}}{\sim} N(0, \sigma_\alpha^2)$. In some applications, specifically in epidemiological studies, it usually happens that after certain times lag, shape of the distribution of subjects' responses will change. Therefore, it is usually important that different subjects can be categorized based on behavior of their responses over time. One of the most fundamental reformations happens by shifting from a symmetric to an asymmetric distribution, or vice versa. We now propose $\varepsilon_{it} \stackrel{\text{iid}}{\sim} SN(0, \sigma_\varepsilon^2; \delta_{it})$. The parameter δ_{it} is the skewness parameter which we propose being δ for $t < c_i$, specifying a skewed distribution and is set to zero for $t > c_i$ which determines the normal distribution. Using the hierarchical representation of this distribution, we have $Y_{it}|z_{it}, \alpha_i \stackrel{\text{iid}}{\sim} N(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i + \delta_{it}z_{it}, \sigma^2)$ where $z_{it} \stackrel{\text{iid}}{\sim} N(0, 1)I\{z_{it} > 0\}$. To have a flexible modeling structure along with being able to cluster subjects, we consider the DP as a prior for unknown distribution of c_i 's. More specifically, we assume $c_i|G \stackrel{\text{iid}}{\sim} G$ where $G \sim DP(M, G_0)$ and G_0 is a discrete uniform distribution on $\{1, \dots, T\}$. For simplicity, we assume α_i and c_i be independent.

3.1 The initial conditions problem

In fitting a transition model, the individual effects that capture the unobserved heterogeneity are usually correlated with the initial state y_{i0} . Ignoring this correlation, in the case of fixed T and large n , produces seriously biased parameter estimates, known as the initial conditions problem (Kazemi and Davies (2002)). A pragmatic solution is to consider the following equation for the y_{i0} and then jointly model it with the subsequent responses. More specifically, we consider

$$Y_{i0} = \mathbf{x}'_{i0}\boldsymbol{\beta}_0 + u_{i0}, \quad i = 1, \dots, n, \quad (3.2)$$

where the \mathbf{x}_{i0} 's are initial period covariates. We further assume that the $u_{i0} = \varphi_1\alpha_i + \varphi_2c_i + \varepsilon_{i0}$ where $\varepsilon_{i0} \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$. Combining Equations (3.1) and (3.2), the full model is given by

$$\mathbf{Y}_i = \tilde{\mathbf{X}}_i\boldsymbol{\theta} + \mathbf{w}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i\mathbf{z}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (3.3)$$

where $\mathbf{Y}_i = (Y_{i0}, Y_{i1}, \dots, Y_{iT})'$, $\boldsymbol{\theta} = (\boldsymbol{\beta}'_0, \gamma, \boldsymbol{\beta}')'$, $\boldsymbol{\xi}_i = (\alpha_i, c_i)'$, $\mathbf{w}_1 = (\varphi_1, \mathbf{1}')'$, $\mathbf{w}_2 = (\varphi_2, \mathbf{0}')'$, $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2)$, $\boldsymbol{\delta}_i = (0, \delta\mathbf{1}'_{c_i}, \mathbf{0}'_{T-c_i})'$ and the model matrix $\tilde{\mathbf{X}}_i$ is defined according to $\boldsymbol{\theta}$.

4 Bayesian Estimation

In this section, we use the Gibbs sampler to obtain the parameter estimates. The following distributions are adopted as priors: Inverse gamma priors, $IG(\tau_1, \tau_2)$ for σ_ε^2 and $IG(\delta_1, \delta_2)$ for σ_α^2 , the $N(\boldsymbol{\theta}_0, \boldsymbol{\Lambda})$ for $\boldsymbol{\theta}$, the $N(\varphi_{0j}, \sigma_{\varphi_j}^2)$ for $\varphi_j, j = 1, 2$ and $N(0, \sigma_\delta^2)$ for δ . Data analysis are conducted by using the Gibbs sampler which simulates iteratively from the complete conditional posterior (CCP) distribution of each unknown parameter, given the current values of all other model parameters and the observations. In below, we derive the related CCPs.

Being applicable in the OpenBUGs software, we use an equivalent structure introduced by Ishwaran and James (2001) which considers the following model

$$\begin{aligned} \mathbf{Y}_i | c_{\lambda_i}, \varphi_1, \varphi_2, \alpha_i, \boldsymbol{\theta}, \lambda, \sigma_\varepsilon^2, \delta &\stackrel{\text{iid}}{\sim} N_{T+1} \left(\tilde{\mathbf{X}}_i\boldsymbol{\theta} + \mathbf{w}_1\alpha_i + \mathbf{w}_2c_{\lambda_i} + \boldsymbol{\delta}_{\lambda_i}\mathbf{z}_i, \sigma_\varepsilon^2\mathbf{I}_{T+1} \right), \\ \lambda_i &\stackrel{\text{iid}}{\sim} G(\lambda_i) = \sum_{j=1}^C \pi_j \Delta_j(\lambda_i), \quad \text{where } \pi_j \stackrel{\text{iid}}{\sim} SB(M), \\ c_j &\stackrel{\text{iid}}{\sim} DU\{1, \dots, T\}, \end{aligned} \quad (4.1)$$

where $\mathbf{z}_i \stackrel{\text{iid}}{\sim} N_{T+1}(\mathbf{0}, \mathbf{I})I\{\mathbf{z}_i > 0\}$ and $\alpha_i \stackrel{\text{iid}}{\sim} N(0, \sigma_\alpha^2)$. Then, after some algebra is done, we have

$$\boldsymbol{\theta} | \sigma_\varepsilon^2, \varphi_1, \varphi_2, \delta, \mathbf{z}, \boldsymbol{\xi}, \mathbf{y} \sim N_p(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta), \quad (4.2)$$

with the mean vector and the covariance matrix

$$\boldsymbol{\mu}_\theta = \boldsymbol{\Sigma}_\theta \left(\boldsymbol{\theta}'_0 \boldsymbol{\Lambda}^{-1} + \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{w}\boldsymbol{\xi}_i - \boldsymbol{\delta}_i\mathbf{z}_i)' \tilde{\mathbf{X}}_i \right), \quad \boldsymbol{\Sigma}_\theta = \left(\boldsymbol{\Lambda}^{-1} + \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^n \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right)^{-1}. \quad (4.3)$$

For the residual variance, we have

$$\sigma_\varepsilon^2 \mid \boldsymbol{\theta}, \boldsymbol{\xi}_i, \varphi_1, \varphi_2, \mathbf{z}_i, \delta, \mathbf{y} \sim IG(\tau_1^*, \tau_2^*), \tag{4.4}$$

where $\tau_1^* = \tau_1 + n(T + 1)/2$ and $\tau_2^* = \tau_2 + \frac{1}{2} \sum_{i=1}^n (\mathbf{r}_i - \mathbf{w}\boldsymbol{\xi}_i - \boldsymbol{\delta}_i \mathbf{z}_i)' (\mathbf{r}_i - \mathbf{w}\boldsymbol{\xi}_i - \boldsymbol{\delta}_i \mathbf{z}_i)$ where $\mathbf{r}_i = \mathbf{Y}_i - \tilde{\mathbf{X}}_i \boldsymbol{\theta}$. Further, for the parameter $\varphi_j, j = 1, 2$, we show that

$$\varphi_j \mid \sigma_\varepsilon^2, \boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{y} \sim N(\mu_{\varphi_j}, \sigma_{\varphi_j}^{2*}), \tag{4.5}$$

where $\mu_{\varphi_j} = \sigma_{\varphi_j}^{2*} \left(\frac{1}{\sigma_{\varphi_j}^2} \varphi_{0j} + \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^n r_{i0} \xi_{ji} \right)$, $\sigma_{\varphi_j}^{2*} = \left(\frac{1}{\sigma_{\varphi_j}^2} + \frac{1}{\sigma_\varepsilon^2} SS_{A_j} \right)^{-1}$, the r_{i0} 's are fitted residuals for $t = 0$ and $SS_{A_j} = \sum_{i=1}^n \xi_{ji}^2$. For the CCP of each random intercept, we derive

$$\alpha_i \mid \sigma_\varepsilon^2, \sigma_\alpha^2, \boldsymbol{\theta}, \mathbf{z}_i, \varphi_1, \mathbf{y} \sim N(\mu_i^*, \sigma_\alpha^{2*}), \tag{4.6}$$

where $\mu_i^* = \boldsymbol{\psi} \mathbf{w}'_1 (\mathbf{r}_i - \mathbf{w}_2 c_{\lambda_i} - \boldsymbol{\delta}_{\lambda_i} \mathbf{z}_i)$ and $\sigma_\alpha^{2*} = \sigma_\varepsilon^2 \psi$ for $\psi = \sigma_\alpha^2 / \sigma_c^2$ with $\sigma_c^2 = \sigma_\varepsilon^2 + (T + \varphi_1^2) \sigma_\alpha^2$. Moreover, for σ_α^2 we obtain

$$\sigma_\alpha^2 \mid \boldsymbol{\xi}, \mathbf{y} \sim IG(\delta_1^*, \delta_2^*) \tag{4.7}$$

where $\delta_1^* = \delta_1 + \frac{n}{2}$ and $\delta_2^* = \delta_2 + \frac{1}{2} SS_{A_1}$. For each λ_i we derive

$$\lambda_i \mid \boldsymbol{\theta}, \sigma_\varepsilon^2, \varphi_1, \varphi_2, \alpha_i, c_{\lambda_i}, \boldsymbol{\pi}, \mathbf{y} \sim \sum_{j=1}^C \pi_j \varphi_T \left(\mathbf{y}_i \mid \tilde{\mathbf{X}}_i \boldsymbol{\theta} + \mathbf{w}_1 \alpha_i + \mathbf{w}_2 c_{\lambda_i}, \sigma_\varepsilon^2 \mathbf{I}_T \right) \Delta_j(\lambda_i), \tag{4.8}$$

For each random effect, let λ_r^* 's, $r = 1, \dots, m$ be m unique values of λ_i 's. We derive

$$f(c_{\lambda_r}^* \mid \boldsymbol{\xi}, \sigma_\varepsilon^2, \varphi_1, \boldsymbol{\theta}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} M_r \varphi_2^2 c_{\lambda_r}^{*2} + \frac{1}{\sigma_\varepsilon^2} c_{\lambda_r}^* \varphi_2 \sum_{\{i: \lambda_i = \lambda_r^*\}} (r_{i0} - \varphi_1 \boldsymbol{\xi}_{1i}) \right\} \text{ for } c_{\lambda_r}^* \in \{1, \dots, T\}, \tag{4.9}$$

where M_r is the number of λ_i 's that equals λ_r^* . We simulate from this CCP by using the Metropolis-Hastings algorithm. By the conjugacy of the multinomial and the Dirichlet distributions, we have

$$\pi_j = \gamma_j^* \prod_{r=1}^{j-1} (1 - \gamma_r^*), \quad j = 1, \dots, C - 1 \tag{4.10}$$

where $\pi_1 = \gamma_1^*$ and $\gamma_j^* \mid \boldsymbol{\lambda}, \mathbf{y} \stackrel{\text{iid}}{\sim} \text{Beta}(1 + M_j, M + \sum_{r=j+1}^C M_r)$. For parameter δ , we have

$$\delta \mid \boldsymbol{\theta}, \varphi_1, \varphi_2, \boldsymbol{\xi}, \sigma_\varepsilon^2, \mathbf{z}, \mathbf{y} \sim N(\mu_\delta^*, \sigma_\delta^{2*}), \tag{4.11}$$

$$\mu_\delta^* = \frac{\sigma_\delta^{2*}}{\sigma_\varepsilon^2} \sum_{i=1}^n \left(\mathbf{y}_{i_{c_i}} - \tilde{\mathbf{X}}_{i_{c_i}} \boldsymbol{\theta} - \mathbf{w}_{c_i} \boldsymbol{\xi}_i \right)' \mathbf{z}_{i_{c_i}}, \quad \sigma_\delta^{2*} = \left(\frac{1}{\sigma_\delta^2} + \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^n \mathbf{z}'_{i_{c_i}} \mathbf{z}_{i_{c_i}} \right)^{-1}, \tag{4.12}$$

where each matrix with index c_i includes the 2nd to the $(c_i + 1)$ -th rows of corresponding matrix. Now, the Gibbs sampler simulates iteratively from these CCPs until converges to stationary distributions. Then, average of samples for each parameter can be used as its Bayes estimate.

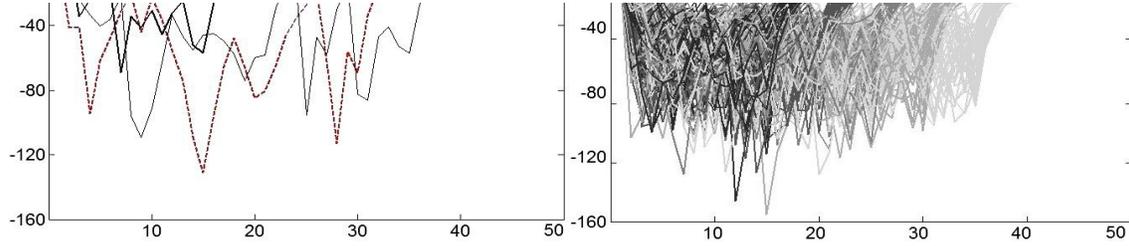


Figure 1: Profiles of generated data points for three subjects (left panel) and for all subjects (right panel).

5 Simulation study

The data generating process is organized to the mixed model

$$y_{it} = \beta_0 + \beta_1 x_{it} + \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, 250, \quad t = 1, \dots, 50. \tag{5.1}$$

We set $\beta_0 = 0$, $\beta_1 = 1$ and $\gamma = 0.5$. The variables $x_{it} \stackrel{iid}{\sim} N(0, \sigma_x^2)$, $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$ and $\varepsilon_{it} \stackrel{ind}{\sim} SN(0, \sigma_\varepsilon^2; \delta_{it})$. The parameters σ_x^2 , σ_α^2 and σ_ε^2 are all set to one. For each subject, $\delta_{it} = -35$ for $t < c_i$ and is set to zero otherwise. For each subject, c_i is randomly generated form the set $\{15, 20, 25, 30, 35\}$ and therefore, each subject is randomly assigned to a cluster. Model (3.2) is considered for generating y_{i0} 's with $\varphi_1 = 0.98$ and $\varphi_2 = 0$. Configuration of generated data points for three randomly selected subjects and for all subjects are depicted in panels of Figure 1.

We fit the mixed model (3.3) by $c_i | G \stackrel{iid}{\sim} G$, $G | M, G_0 \sim DP(M, G_0)$ where G_0 is considered the discrete uniform distribution on $\{1, \dots, 50\}$, Model M1. As a competitive model, we assume model (3.3) with $c_i \stackrel{iid}{\sim} \{1, \dots, 50\}$, Model M2. To implement the Gibbs sampler, the following independent priors are adopted: $N(0, 100)$ for each regression coefficient, φ_1 and φ_2 , $U(0, 1)$ for γ and $IG(0.01, 0.01)$ for variance components. We use the OpenBUGs software with 15000 samples generated after 3500

Table 1: Bayesian estimation results of Models M1 and M2 for the simulated data set.

Model	β_0	β_1	γ	σ_ε^2	δ	σ_α^2	β_{00}	β_{01}	φ_1	φ_2
M1	0.089 (0.064)	1.011 (0.013)	0.501 (0.001)	1.068 (0.020)	-34.270 (0.301)	1.019 (0.096)	-0.383 (1.902)	1.004 (0.087)	-0.148 (0.359)	0.018 (0.049)
M2	0.151 (0.070)	1.012 (0.014)	0.501 (0.001)	1.058 (0.020)	-33.910 (0.323)	1.029 (0.100)	0.548 (0.169)	1.002 (0.064)	0.948 (0.071)	-0.016 (0.005)

* Bayesian standard deviations are given in parentheses.

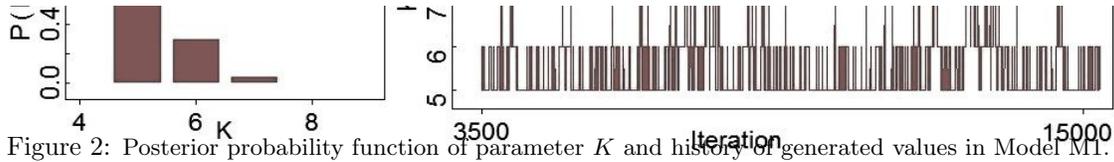


Figure 2: Posterior probability function of parameter K and history of generated values in Model M1.

Table 2: Bayesian estimates of change points of Model M2 and sizes of each cluster are reported.

	c_1^*	c_2^*	c_3^*	c_4^*	c_5^*	c_6^*	c_7^*	c_8^*
c_j^*	15.50	20.50	25.51	30.43	34.93	37.47	40.99	44.97
Size of clusters	56.00	40.00	46.96	44.62	50.18	11.37	0.86	0.01

burn-in. Results, after convergence is achieved, are reported in Table 1. We set the scale parameter $M = 1$ and the truncation parameter $C = 8$. To show that $C = 8$ suffices to approximate infinite mixture, the number of active clusters, K , is also estimated. Estimate of median of K is obtained 5 in Model M1. The posterior probability density function of parameter K and history of generated values in Model M1 are shown in Figure 2. These graphs indicate that we have achieved good approximation of the DP by the use of truncated stick-breaking representation. This is also confirmed by the estimated values of change points and the estimated size of each cluster, reported in Table 2. As is seen, clusters corresponding to the last three change points are inactive (empty) in most runs of Gibbs sampler.

As is seen, estimates of longitudinal effects β_1 and δ are slightly more precise in Model M1 than Model M2. Moreover, Model M1 can be used for the clustering purpose. We cluster the data set by using the estimated values. In each model, the likelihood of each data point is computed for active clusters 1 to 5. Then, each data point is assigned to the cluster with larger likelihood. The ratio of true clustered subjects in Model M1 is obtained 0.9960.

Conclusion

In this paper, we extended the application of DP priors for clustering of longitudinal data. This was specifically done by incorporating the DP as the prior of unknown distribution of the random change point in the structure of the piecewise transition mixed-effects model. It was shown by the simulation study that the ratio of true clustered subjects is high in the proposed model.

References

- Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230.
- Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of American Statistical Association*, **96**(453), 161–173.
- Kazemi, I. and Davies, R.B. (2002). The asymptotic bias of MLEs for dynamic panel data models, in: Stasinopoulos, M. and Touloumi, G. (Eds.), *Proceedings of the 17th International Workshop on Statistical Modelling*, Chania, Greece, pp. 391–395.
- McLain, A.C. and Albert, P.S. (2014). Modeling longitudinal data with a random change point and no time-zero: applications to inference and prediction of the labor curve. *Biometrics*, **70**, 1052–1060.
- Sahu, S.K., Dey, D.K. and Branco, M.D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, **31**, 129–150.



Hill-Type Estimators and Invariance

Mehdi Shams^{1,2}, Gholamreza Hesamian²

¹Department of Mathematical Sciences, University of Kashan, Isfahan, Iran.

²Department of Statistics, Payame Noor University, Tehran 19395-3697, Iran.

Abstract: The Hill-type estimators has been used in extreme value theory in order to estimate the tail index related to a distribution function with a positive index. We will introduce a new Hill-type estimator which is invariant. In addition, its asymptotic distributional representation and asymptotic normality are investigated.

Keywords Invariant Hill-type, heavy tailed index estimator, second order regular variation, extreme value theory, regular variation.

Mathematics Subject Classification (2010): 62F12 62G30 62G32.

1 Introduction

In this paper we study the asymptotic properties of a class unbiased location invariant Hill-type heavy tailed index estimators. Let $\{X_n, n \geq 1\}$ be an i.i.d sequence with the distribution function F belonging to the domain of a attraction of an extreme value distribution G_γ , ($F \in D(G_\gamma)$), with

$$G_\gamma(z) = \begin{cases} \exp(-(1 + \gamma z)^{\frac{-1}{\gamma}}) & \text{for } 1 + \gamma z > 0, & \text{if } \gamma \neq 0, \\ \exp(-\exp(-z)) & \text{for } z \in R, & \text{if } \gamma = 0. \end{cases}$$

i.e., there exist normalizing constant $a_n > 0$ and $b_n \in R$ s.t. $F^n(a_n z + b_n) \rightarrow G_\gamma(z)$, as $n \rightarrow \infty$ for all $z \in R$. We denote by γ the tail index and G_γ is the generalized extreme value distribution (*GEV*). Gendenko (1943) has established that G_γ represents the only possible limit distribution of the maximum $X_{(n,n)}$ suitably normalized, where $X_{(1,n)} \leq X_{(2,n)} \leq \dots \leq X_{(n,n)}$ are the order statistics of X_1, X_2, \dots, X_n . Here, the shape parameter γ plays a central rule, in sense that it rules the tail weight of F . The problem of estimating the tail index γ has received much attention by many researchers in statistics of extremes. In this regards, several estimators have been suggested for tail index γ . For instance, a set of statistical procedures based on extreme value theory results can be found in Reiss and Thomas (2001). Other recent surveys on this subject are Resnick (1997) and Csörgö and Viharos (1998). The applications of the heavy tailed distribution may be found in many fields such

²Speaker's Name: Mehdi Shams, mehdishams@kashanu.ac.ir

as insurance, finance, climatology and environmental science (Emberechets et al. (1997)) and the problem of estimating the heavy tailed index has been studied extensively. For positive γ , Hill (1975) introduced the well known Hill estimator, which is $\hat{\gamma}_n^H(k) = \frac{1}{k} \sum_{i=1}^{k-1} \ln X_{(n-i,n)} - \ln X_{(n-k,n)}$. $\hat{\gamma}_n^H$ has been largely used for the estimation of the shape tail index parameter γ , whenever the underlying distribution is in the domain of attraction of the Frechet distributions, which can be regarded as a particular case of *GEV*-type distribution with the possible inclusion of location and scale parameters $G_\gamma(x; \lambda, \delta) = \exp(-(1+\gamma(x-\lambda)/\delta)^{-\frac{1}{\gamma}})$ for $(1+\gamma(x-\lambda)/\delta) > 0$ and $\gamma > 0$. Here, as a usual, the location parameter λ is real and the scale parameter δ is positive. For general γ , the exact common estimator are Pickands estimator (Pickands, 1975) and moment estimator (Dekkers et. al (1989)) defined as $\hat{\gamma}_n^P(k) = \ln \frac{X_{(n,n-[k/4])} - X_{(n,n-[k/2])}}{X_{(n,n-[k/2])} - X_{(n,n-k)}}$, $\hat{\gamma}_n^M(k) = \hat{\gamma}_n^+ + \hat{\gamma}_n^-$ respectively, where $\hat{\gamma}_n^+ = M_n^{(1)}$, $\hat{\gamma}_n^- = 1 - \frac{1}{2} \{1 - (M_n^{(1)})^2 / M_n^{(2)}\}^{-1}$, $M_n^{(j)} = \frac{1}{k} \sum_{i=0}^{k-1} (\ln X_{(n-i,n)} - \ln X_{(n-k,n)})^j$, $j > 0$. Note that the moment estimator $\hat{\gamma}_n^M$ is consistent for any real γ . This is one of the most used estimators for γ , if there is no a priori knowledge about the tail parameters. In many applications, both Pickands and moment estimators are not invariant to the affine transformation. It is mentioned that the Pickands estimator is location invariant. In a general case, Fraga Alves (2001) established the scale and location invariant Hill-type estimator given by $\hat{\gamma}_n^H(k_0, k) = \frac{1}{k_0} \sum_{i=0}^{k_0-1} \ln \frac{X_{(n-i,n)} - X_{(n-k,n)}}{X_{(n-k_0,n)} - X_{(n-k,n)}}$, where

$$k = k_n = o(n), k_0 = o(k_n), k_n \rightarrow \infty, k_0 \rightarrow \infty, \text{ as } n \rightarrow \infty. \quad (1)$$

It is easy to check that $\lim_{n \rightarrow \infty} \hat{\gamma}_n^H(k_0, k) = \gamma$, in probability and also $\sqrt{k_0}(\hat{\gamma}_n^H(k_0, k) - \gamma) \rightarrow^d Z$, where Z is normally distributed with mean zero and variance γ^2 . The asymptotic properties of $\hat{\gamma}_n(k_0, k)$ have been considered when $\frac{U(tx)/U(t) - x^\gamma}{a(t)} \rightarrow x^\gamma \frac{x^\rho - 1}{\rho}$, holds as $t \rightarrow \infty$, where $U(t) = F^{\leftarrow}(1 - 1/t)$, $t \geq 1$, where F^{\leftarrow} denotes the inverse function of $F(x)$. Based on the methods provided by Fraga Alves (2001) and Dekkers et al. (1989), Ling et al. (2007a, 2007b) proposed a kind of location invariant moment-type tail index estimator and considered its asymptotic properties under some second order regular varying conditions. Here we are intrusted in the class of semi-parametric heavy tail estimator introduced by Caeiro and Gomes (2002) as $\gamma_n^\alpha(k) = \frac{\Gamma(\alpha)}{M_n^{\alpha-1}(k)} (M_n^{2\alpha}(k) / \Gamma(2\alpha+1))^{1/2}$ where $\alpha \geq 1$ and $\Gamma(\cdot)$ is the gamma function. They considered the asymptotic distributional representation of $M_n^\alpha(k)$ and the choice of tuning parameter α s.t. $\gamma_n^\alpha(k)$ is asymptotically normal with asymptotic null bias under the assumption that $\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} \rightarrow \frac{x^{\rho-1}}{\rho}$, $x > 0$, and $\sqrt{k}A(n/k) \rightarrow \lambda$ which is finite. Motivated by the works of Fraga Alves (2001) and Caeiro and Gomes (2002), we propose a new class of location invariant estimators for a heavy tailed distribution based on the asymptotic distributional representation of the following statistics

$$M_n^\alpha(k_0, k) = \frac{1}{k_0} \sum_{i=1}^{k_0-1} \left(\ln \frac{X_{(n-i,n)} - \ln X_{(n-k,n)}}{X_{(n-k_0,n)} - \ln X_{(n-k,n)}} \right)^\alpha, \alpha > 0.$$

The asymptotic distributional representation of $M_n^\alpha(k_0, k)$ will be derived under the following second order regular variation condition

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^{\gamma-1}}{\gamma}}{A(t)} = \Psi_{\gamma, \rho}(x), \text{ for all } x > 0, \tag{2}$$

where

$$\Psi_{\gamma, \rho}(x) = \begin{cases} \frac{x^{\gamma+\rho}-1}{\gamma+\rho} & \text{if } \gamma + \rho \neq 0, \\ \ln x & \text{if } \gamma + \rho = 0. \end{cases}$$

and $\rho < 0$ is the second order parameter and $|A(t)| \in RV_\rho$ (Corollary 2.3.5 of the de Haan and Ferreira (2006)). Here $f \in RV_\beta$ means $\lim_{t \rightarrow \infty} f(tx)/f(t) = x^\beta$ for all $x > 0$. Now based on the convergence $M_n^\alpha(k_0, k) \rightarrow^P \Gamma(\alpha+1)\gamma^\alpha$ as $k_0 \rightarrow \infty, k_0 = o(k)$, a location invariant Hill-type estimator for the heavy tailed index may be defined by

$$\hat{\gamma}_n^\alpha(k_0, k) = \frac{\Gamma(\alpha)}{M_n^{\alpha-1}(k_0, k)} \left(\frac{M_n^{2\alpha}(k_0, k)}{\Gamma(2\alpha + 1)} \right)^{1/2}, \alpha \geq 1, \tag{3}$$

which converges to γ in probability.

In this paper, we investigate the asymptotic distributional representation of $M_n^\alpha(k_0, k), \hat{\gamma}_n^\alpha(k_0, k)$, and the optimal choice of the sample fraction k_0 by MSE for some special distributions.

2 The main results

For the heavy tail distribution, i.e. $\gamma > 0$, we know that $F \in D(G_\gamma) \Leftrightarrow 1 - F \in RV_{-1/\gamma} \Leftrightarrow U \in RV_\gamma$. i.e. $U = (1/(1 - F))^\leftarrow$ is regularly varying with index γ which is equivalent to say that there exist a function $a(t) > 0$ s.t. $\frac{U(tx) - U(t)}{a(t)} \rightarrow \frac{x^\gamma - 1}{\gamma}$ for $x > 0$ (de Haan, 1984). To do this, consider the following notations to simplify the statement

$$\sigma_\alpha = \sqrt{\Gamma(2\alpha + 1) - \Gamma^2(\alpha + 1)}, \quad \mu_\alpha(\rho) = \frac{\Gamma(\alpha)}{\rho} \frac{1 - (1 - \rho)^\alpha}{(1 - \rho)^\alpha},$$

$$V_\alpha = \frac{1}{4} \left\{ \frac{\Gamma(4\alpha)}{\alpha\Gamma^2(2\alpha)} + \frac{4\Gamma(2\alpha - 1)}{\Gamma^2(\alpha)} - \frac{2\Gamma(3\alpha)}{\alpha\Gamma(\alpha)\Gamma(2\alpha)} - 1 \right\}, \tag{4}$$

and

$$b_\alpha(\gamma) = (1 + \gamma)^{1-\alpha} - \frac{1}{2}(1 + \gamma)^{-2\alpha} - \frac{1}{2}. \tag{5}$$

The first result is about the asymptotic distributional representation of $M_n^\alpha(k_0, k)$.

Theorem 2.1. *Suppose that (1) holds for $\gamma > 0$, and the intermediate k_0 and k satisfy (1). Then $M_n^\alpha(k_0, k)$ converges in probability to $\Gamma(\alpha + 1)/\gamma^\alpha$. Furthermore, if the second order framework in (2) holds, we have the following asymptotic distributional representation*

$$M_n^\alpha(k_0, k) =^d \gamma^\alpha \Gamma(1 + \alpha) + \frac{\gamma^\alpha \sigma_\alpha}{\sqrt{k_0}} P_n^{(\alpha)} + \alpha \gamma^\alpha \mu_\alpha(-\gamma) \left(\frac{k_0}{k}\right)^\gamma (1 + o_p(1)) + B_n,$$

where $P_n^{(\alpha)}$ is an asymptotically standard normal r.v., and

$$B_n = \begin{cases} \frac{\alpha \gamma^\alpha \rho \mu_\alpha(\rho)}{\gamma + \rho} A(n/k) \left(\frac{k_0}{k}\right)^{-\rho} (1 + o_p(1)) & \text{if } \gamma + \rho \neq 0, \\ \alpha \gamma^{\alpha+1} \mu_\alpha(-\gamma) A(n/k) \left(\frac{k_0}{k}\right)^\gamma \ln\left(\frac{k_0}{k}\right) (1 + o_p(1)) & \text{if } \gamma + \rho = 0. \end{cases}$$

Proof: For the asymptotic distributional representation of $M_n^\alpha(k_0, k)$, we only consider the case of $\gamma + \rho \neq 0$, so

$$\begin{aligned} \ln \frac{U(tx) - U(t)}{U(ty) - U(t)} &= \ln \frac{x^\gamma - 1}{y^\gamma - 1} + \frac{\gamma}{\gamma + \rho} \left[\frac{x^{\gamma+\rho} - 1}{x^\gamma - 1} - \frac{y^{\gamma+\rho-1} - 1}{y^\gamma - 1} \right] A(t) (1 + o(1)) \\ &= \gamma \ln \frac{x}{y} + \ln\left(1 + \frac{y^{-\gamma} - x^{-\gamma}}{1 - y^{-\gamma}}\right) + \frac{\gamma}{\gamma + \rho} \left[\frac{x^{-\rho} - x^{-\gamma}}{1 - x^{-\gamma}} - \frac{y^\rho - y^{-\gamma}}{1 - y^{-\gamma}} \right] A(t) (1 + o(1)). \end{aligned}$$

Let Y_1, Y_2, \dots, Y_n be i.i.d Pareto r.v.s with $F_Y(y) = 1 - 1/y, y \geq 1$. Now replace t by $Y_{(n-k,n)}$, x by $Y_{(n-i,n)}/Y_{(n-k,n)}$ and y by $Y_{(n-k_0,n)}/Y_{(n-k,n)}$, respectively, and note that $\{X_i\}_{i=1}^n =^d \{U(Y_i)\}_{i=1}^n$, $(Y_{(n-i,n)}/Y_{(n-k,n)})^{-\gamma} < (Y_{(n-k_0,n)}/Y_{(n-k,n)})^{-\gamma} \rightarrow 0$ in probability uniformly for $i = 0, 1, \dots, k_0 - 1$, and $\{Y_{(n-i,n)}/Y_{(n-k_0,n)}\}_{i=1}^{k_0-1} =^d \{Y_{(k_0-i,k_0)}\}_{i=1}^{k_0-1}$. So,

$$\begin{aligned} \ln \frac{X_{(n-i,n)} - X_{(n-k,n)}}{X_{(n-k_0,n)} - X_{(n-k,n)}} &=^d \ln \frac{U(Y_{(n-i,n)}) - U(Y_{(n-k,n)})}{U(Y_{(n-k_0,n)}) - U(Y_{(n-k,n)})} = \\ &\gamma \ln \frac{Y_{(n-i,n)}}{Y_{(n-k_0,n)}} + \left(\frac{Y_{(n-k_0,n)}}{Y_{(n-k,n)}}\right)^{-\gamma} \left[1 - \left(\frac{Y_{(n-i,n)}}{Y_{(n-k_0,n)}}\right)^{-\gamma}\right] (1 + o_p(1)) \\ &+ \frac{\gamma}{\gamma + \rho} \left[\left(\frac{Y_{(n-k_0,n)}}{Y_{(n-k,n)}}\right)^{-\gamma} \left(1 - \left(\frac{Y_{(n-i,n)}}{Y_{(n-k_0,n)}}\right)^{-\gamma}\right) - \left(\frac{Y_{(n-k_0,n)}}{Y_{(n-k,n)}}\right)^\rho \left(1 - \left(\frac{Y_{(n-i,n)}}{Y_{(n-k_0,n)}}\right)^\rho\right) \right] \\ &\times A(Y_{(n-k,n)}) (1 + o_p(1)) =^d \gamma \ln Y_{(k_0-i,k_0)} + [1 - Y_{(k_0-i,k_0)}^{-\gamma}] \left(\frac{k_0}{k}\right)^\gamma (1 + o_p(1)) \\ &+ \frac{\gamma}{\gamma + \rho} \left[(1 - Y_{(k_0-i,k_0)}^{-\gamma}) \left(\frac{k_0}{k}\right)^\gamma - (1 - Y_{(k_0-i,k_0)}^\rho) \left(\frac{k_0}{k}\right)^{-\rho} \right] A(n/k) (1 + o_p(1)). \end{aligned}$$

By Taylor's expansion, we may get

$$\begin{aligned} M_n^\alpha(k_0, k) &=^d \frac{1}{k_0} \sum_{i=1}^{k_0-1} \left\{ (\gamma \ln Y_{(k_0-i,k_0)})^\alpha + \alpha \gamma^\alpha (\ln Y_{(k_0-i,k_0)})^{\alpha-1} \frac{1 - Y_{(k_0-i,k_0)}^{-\gamma}}{\gamma} \right. \\ &\times \left(\frac{k_0}{k}\right)^\gamma (1 + o_p(1)) + \alpha \frac{\gamma^\alpha}{\gamma + \rho} \ln Y_{(k_0-i,k_0)}^{\alpha-1} \left[(1 - Y_{(k_0-i,k_0)}^{-\gamma}) \left(\frac{k_0}{k}\right)^\gamma - 1 - Y_{(k_0-i,k_0)}^\rho \right. \\ &\left. \left.\left(\frac{k_0}{k}\right)^{-\rho} \right] \times A(n/k) (1 + o_p(1)) \right\} + A^2(n/k) O_p\left(\left(\frac{k_0}{k}\right)^{2\gamma} + \left(\frac{k_0}{k}\right)^{-2\rho} + \left(\frac{k_0}{k}\right)^{\gamma-\rho}\right). \end{aligned}$$

Note that $E[(\ln Y_1)^\alpha] = \Gamma(\alpha + 1)$, $Var[(\ln Y_1)^\alpha] = \sigma_\alpha$, and $E[(\ln Y_1)^{\alpha-1}(Y_1^\rho - 1)/\rho] = \mu_\alpha(\rho)$. So,

$$P_n^{(\alpha)} = \frac{\frac{1}{k_0} \sum_{i=1}^{k_0} (\ln Y_i)^\alpha - \Gamma(\alpha + 1)}{\sigma_\alpha / \sqrt{k_0}} \rightarrow^d N(0, 1),$$

and then by the law of large numbers, the asymptotic distributional representation of $M_n^\alpha(k_0, k)$ is

$$M_n^\alpha(k_0, k) =^d \gamma^\alpha \Gamma(1 + \alpha) + \frac{\gamma^\alpha \sigma_\alpha}{\sqrt{k_0}} P_n^{(\alpha)} + \alpha \gamma^\alpha \mu_\alpha(-\gamma) \left(\frac{k_0}{k}\right)^\gamma (1 + o_p(1)) + \frac{\alpha \rho \gamma^\alpha}{\gamma + \rho} \mu_\alpha(\rho) A(n/k) \left(\frac{k_0}{k}\right)^{-\rho} (1 + o_p(1)),$$

which is the desired result.

Based on Theorem 2.1, the asymptotic distributional representation of the proposed estimator $\widehat{\gamma}_n^\alpha(k_0, k)$ in (3), which is the following result.

Theorem 2.2. *Under the condition of theorem 2.1, the following asymptotic distributional representation*

$$\widehat{\gamma}_n^\alpha(k_0, k) =^d \gamma + \frac{\gamma \sqrt{V_\alpha}}{\sqrt{k_0}} T_n^{(\alpha)} + b_\alpha(\gamma) \left(\frac{k_0}{k}\right)^\gamma + o_p\left(\frac{1}{\sqrt{k_0}}\right) + o_p\left(\left(\frac{k_0}{k}\right)^\gamma\right) + R_n,$$

holds, where $T_n^{(\alpha)}$ is asymptotically standard normal and

$$R_n = \begin{cases} \frac{-\gamma}{\gamma + \rho} b_\alpha(-\rho) A(n/k) \left(\frac{k_0}{k}\right)^{-\rho} (1 + o_p(1)) & \text{if } \gamma + \rho \neq 0, \\ \gamma b_\alpha(\gamma) A(n/k) \left(\frac{k_0}{k}\right)^\gamma \ln\left(\frac{k_0}{k}\right) (1 + o_p(1)) & \text{if } \gamma + \rho = 0. \end{cases}$$

Furthermore, if there exist finite λ_1 , and λ_2 s.t. $\sqrt{k_0} \left(\frac{k_0}{k}\right)^\gamma \rightarrow \lambda_1$ and $\sqrt{k_0} A(n/k) \rightarrow \lambda_2$, then $\sqrt{k_0}(\widehat{\gamma}_n^\alpha(k_0, k) - \gamma) \rightarrow N(\lambda_1 b_\alpha(\gamma), \gamma^2 V_\alpha)$, where V_α and $b_\alpha(\gamma)$ are defined in (3) and (5) respectively. Consequently for every $\gamma > 0$, there exists α_0 given by $\alpha_0 = \alpha_0(\gamma) = \ln(1 + \gamma + (1 + \gamma)^2 - 1) / \ln(1 + \gamma)$, s.t. $b_{\alpha_0}(\gamma) = 0$, i.e. $\widehat{\gamma}_n^{\alpha_0}(k_0, k)$ has asymptotic null bias, even when $\sqrt{k_0} \left(\frac{k_0}{k}\right)^\gamma \rightarrow \lambda_1 \neq 0$ and $\sqrt{k_0} A(n/k) \rightarrow \lambda_2 \neq 0$.

Proof: Note that $\{X_i\}_{i=1}^n =^d \{U(Y_i)\}_{i=1}^n$ and $\{Y_{(n-i,n)}/Y_{(n-k_0,n)}\}_{i=1}^{k_0-1} =^d \{Y_{(k_0-i,k_0)}\}_{i=1}^{k_0-1}$. For $\gamma + \rho \neq 0$, by using the conditions of theorem, following result

$$\left(\frac{M_n^{2\alpha}(k_0, k)}{\Gamma(2\alpha + 1)}\right)^{1/2}, \frac{\Gamma(\alpha)}{M_n^{\alpha-1}(k_0, k)} =^d (W_n, Z_n), \tag{6}$$

holds, where

$$W_n = \gamma^\alpha \left[1 + \left(\frac{\sigma_{2\alpha}}{2\Gamma(2\alpha+1)} \frac{P_n^{(2\alpha)}}{\sqrt{k_0}} + \frac{\mu_{2\alpha}(-\gamma)}{2\Gamma(2\alpha)} \left(\frac{k_0}{k}\right)^\gamma + \frac{\rho \mu_{2\alpha}(\rho)}{2(\gamma+\rho)\Gamma(2\alpha)} A\left(\frac{n}{k}\right) \left(\frac{k_0}{k}\right)^{-\rho} (1 + o_p(1)) \right], \right.$$

and

$$Z_n = \gamma^{1-\alpha} \left[1 - \left(\frac{\sigma_{\alpha-1}}{\Gamma(\alpha)} \frac{P_n^{(\alpha-1)}}{\sqrt{k_0}} + \frac{\mu_{\alpha-1}(-\gamma)}{\Gamma(\alpha-1)} \left(\frac{k_0}{k} \right)^\gamma + \frac{\rho \mu_{\alpha-1}(\rho)}{2(\gamma+\rho)\Gamma(\alpha-1)} A\left(\frac{n}{k}\right) \left(\frac{k_0}{k} \right)^{-\rho} \right) (1 + o_p(1)) \right].$$

Note that (6) may be deduced from Delta-method, as for arbitrary $a, b \in R$, we have

$$\begin{aligned} & a(M_n^{2\alpha}(k_0, k)/\Gamma(2\alpha + 1))^{\frac{1}{2}} + b\Gamma(\alpha)(M_n^{\alpha-1}(k_0, k))^{-1} =^d \\ & a \left[\frac{1}{\Gamma(2\alpha+1)} \frac{1}{k_0} \sum_{i=0}^{k_0-1} \left(\ln \frac{U(Y_{(n-i,n)}) - U(Y_{(n-k,n)})}{U(Y_{(n-k_0,n)}) - U(Y_{(n-k,n)})} \right) 2\alpha \right]^{\frac{1}{2}} \\ & + b\Gamma(\alpha) \left[\frac{1}{k_0} \sum_{i=0}^{k_0-1} \left(\ln \frac{U(Y_{(n-i,n)}) - U(Y_{(n-k,n)})}{U(Y_{(n-k_0,n)}) - U(Y_{(n-k,n)})} \right) 2\alpha \right]^{-1} =^d \\ & aW_n + b\gamma^{1-\alpha} \frac{1}{1-(1-Z_n)\gamma^{\alpha-1}} = aW_n + bZ_n. \end{aligned}$$

Hence

$$\begin{aligned} \widehat{\gamma}_n^\alpha(k_0, k) & =^d Z_n W_n = \gamma \left[1 + \frac{1}{k_0} \frac{\sigma_{2\alpha}}{2\Gamma(2\alpha+1)} P_n^{(2\alpha)} - \frac{\sigma_{\alpha-1}}{\Gamma(\alpha)} P_n^{(\alpha-1)} \right. \\ & + \left. \left(\frac{\mu_{2\alpha}(-\gamma)}{2\Gamma(2\alpha)} - \frac{\mu_{\alpha-1}(-\gamma)}{\Gamma(\alpha-1)} \right) \left(\frac{k_0}{k} \right)^\gamma + \frac{\rho}{\rho+\gamma} \left(\frac{\mu_{2\alpha}(\rho)}{2\Gamma(2\alpha)} - \frac{\mu_{\alpha-1}(\rho)}{\Gamma(\alpha-1)} \right) \right. \\ & \left. A\left(\frac{n}{k}\right) \left(\frac{k_0}{k} \right)^{-\rho} (1 + o_p(1)) \right] + o_p\left(\frac{1}{\sqrt{k_0}}\right) + o_p\left(\left(\frac{k_0}{k}\right)^\gamma\right). \end{aligned}$$

Now consider

$$Q_n^{(\alpha)} = \frac{\sigma_{2\alpha}}{2\Gamma(2\alpha + 1)} P_n^{(2\alpha)} - \frac{\sigma_{\alpha-1}}{\Gamma(\alpha)} P_n^{(\alpha-1)}, \quad T_n^{(\alpha)} = \frac{Q_n^{(\alpha)}}{\sqrt{Var(Q_n^{(\alpha)})}}.$$

Let $f_{k_0}(t)$ denote the characteristic function of $Q_n^{(\alpha)}$. Noting the expression of

$$P_n^{(\alpha)} = \frac{1}{k_0} \sum_{i=1}^{k_0} (\ln Y_i)^\alpha - \Gamma(\alpha + 1) \xrightarrow{d} N(0, 1),$$

$\sigma_\alpha / \sqrt{k_0}$

and that of V_n in (4), we have

$$\begin{aligned} f_{k_0}(t) & = E \exp\left\{ it \frac{\sigma_{2\alpha}}{2\Gamma(2\alpha+1)} P_n^{(\alpha)} - it \frac{\sigma_{\alpha-1}}{\Gamma(\alpha)} P_n^{(\alpha-1)} \right\} \\ & = E \exp\left\{ \frac{it}{\sqrt{k_0}} \sum_{j=1}^{k_0} \frac{(\ln Y_j)^{2\alpha} - \Gamma(2\alpha+1)}{2\Gamma(2\alpha+1)} - \frac{it}{\sqrt{k_0}} \sum_{j=1}^{k_0} \frac{(\ln Y_j)^{\alpha-1} - \Gamma(\alpha)}{\Gamma(\alpha)} \right\} \\ & = \prod_{j=1}^{k_0} E \exp\left\{ \frac{it}{\sqrt{k_0}} \left(\frac{(\ln Y_j)^{2\alpha} - \Gamma(2\alpha+1)}{2\Gamma(2\alpha+1)} - \frac{(\ln Y_j)^{\alpha-1} - \Gamma(\alpha)}{\Gamma(\alpha)} \right) \right\} \\ & = \prod_{j=1}^{k_0} \left\{ 1 - \frac{t^2}{2k_0} E \left(\left(\frac{(\ln Y_j)^{2\alpha} - \Gamma(2\alpha+1)}{2\Gamma(2\alpha+1)} - \frac{(\ln Y_j)^{\alpha-1} - \Gamma(\alpha)}{\Gamma(\alpha)} \right) \right)^2 + o\left(\frac{1}{k_0}\right) \right\} \\ & = \left\{ 1 - \frac{t^2}{2k_0} V_\alpha + o\left(\frac{1}{k_0}\right) \right\}^{k_0} \rightarrow \exp(-t^2 V_\alpha / 2). \end{aligned}$$

So, $Q_n^{(\alpha)}$ converges in distribution to a normal r.v. with null bias and variance V_α and $T_n^{(\alpha)}$ is an asymptotically standard normal r.v. Therefore, noting

$$\gamma \left[\frac{\mu_{2\alpha}(-\gamma)}{2\Gamma(2\alpha)} - \frac{\mu_{\alpha-1}(-\gamma)}{\Gamma(\alpha-1)} \right] = b_\alpha(\gamma), \quad \frac{\gamma\rho}{\gamma+\rho} \left[\frac{\mu_{2\alpha}(\rho)}{2\Gamma(2\alpha)} - \frac{\mu_{\alpha-1}(\rho)}{\Gamma(\alpha-1)} \right] = \frac{-\gamma}{\gamma+\rho} b_\alpha(-\rho),$$

we obtain the asymptotic distributional representation of $\widehat{\gamma}_n^{(\alpha)}(k_0, k)$. The remanding part of theorem is immediately proved.

Lemma 2.1. Let $k(n)$ and $k_0(k)$ be intermediate sequences s.t. $k = k(n) \rightarrow \infty$, $k_0 = k_0(k) \rightarrow \infty$, $k/n \rightarrow 0$ and $k_0/k \rightarrow 0$ as $n \rightarrow \infty$. Then $\lim_{n \rightarrow \infty} \hat{\gamma}_n^H(k_0, k) = \gamma$, in probability.

Proof: Let Y_1, Y_2, \dots, Y_n be i.i.d. with distribution function $F_Y(y) = 1 - y^{-1}$, $y \geq 1$. Then $X_{(i,n)} =^d U(Y_{(i,n)})$, for $i = 1, 2, \dots, n$ where $U = (1/(1 - F_Y))^\leftarrow$ is regular varying with index γ . Clearly we have $\{\frac{Y_{(n-i,n)}}{Y_{(n-k_0,n)}}\}_{i=0}^{k_0-1} = \{Y_{(k_0-i,k_0)}\}_{i=0}^{k_0-1}$, where $\ln Y_{(i,n)} =^d E_{(i,n)}$, for $i = 1, 2, \dots, n$, and E_1, E_2, \dots, E_n are i.i.d. with underlying exponential distribution. Also for any intermediate sequence $k = k(n)$, $\lim_{n \rightarrow \infty} (k/n)Y_{(n-k,n)} = 1$ in probability. Hence

$$\begin{aligned} \hat{\gamma}_n^H(k_0, k) &=^d \frac{1}{k_0} \sum_{i=0}^{k_0-1} \ln \frac{U(Y_{(n-i,n)}) - U(Y_{(n-k,n)})}{U(Y_{(n-k_0,n)}) - U(Y_{(n-k,n)})} = \\ &= \frac{1}{k_0} \sum_{i=0}^{k_0-1} \ln \frac{U(Y_{(k-i,k)}Y_{(n-k,n)}) - U(Y_{(n-k,n)})}{U(Y_{(k-k_0,n)}Y_{(n-k,n)}) - U(Y_{(n-k,n)})} = \\ &= \frac{\gamma}{k_0} \sum_{i=0}^{k_0-1} \ln \frac{Y_{(k-i,k)}}{Y_{(k-k_0,k)}} (1 + o_p(1)) = \frac{\gamma}{k_0} \sum_{i=0}^{k_0-1} \ln Y_{(k_0-i,k_0)} (1 + o_p(1)) \\ &= \frac{\gamma}{k_0} \sum_{i=0}^{k_0-1} E_{(k_0-i,k_0)} (1 + o_p(1)) = \frac{\gamma}{k_0} \sum_{i=1}^{k_0} E_i (1 + o_p(1)), \end{aligned}$$

and so $\lim_{n \rightarrow \infty} \hat{\gamma}_n^H(k_0, k) = \gamma$, in probability, as $n \rightarrow \infty$.

3 Conclusion

In this paper we studied a new estimator under semi-parametric approach, which behaves similarly to the Hill's estimator, but it is also invariant to the location transformation. This study was motivated by a previous research (Fraga Alves (1999)) in a parametric setup, where the underlying model was a generalized pareto with the positive tail index. The obtained results, in fact, was a extension of the previous ones for $\rho = -\gamma$.

References

- Caeira, E. and Gomes, M. I. (2002). A class of asymptotically unbiased semi-parametric estimators of the tail index. *Sociedad de Estadística e Investigación Operativa*, **11**, 345-364.
- Csörgő, S. and Viharos, L. (1998). Estimating the tail index. *Asymptotic Methods in Probability and Statistics*, B. Amsterdam, 883-881.
- Dekkers, A.L.M., Binmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme value distribution. *Annals of Statistics*, **4**, 1833-1855.
- Emborechets, P., Kluppelberg, C. and Mikosch, T. (1997). Modeling External Events for Insurance and Finance. Springer, New York.

- Fraga Alves, M.I., (1999). Hill-type estimator, not affected by location. Extended Abstract of Statistical Modeling Extreme Values and Additive Laws, Portugal, 36-39.
- Fraga Alves, M.I., (2001). A location invariant Hill-type estimator. *Extreme*, **4**, 199-217.
- Gnedenko, B.V., (1943), Sur la distribution limite de terme maximum d'une serie aleatoire, *Annals of Mathematics*, **44**, 423-453.
- Haan, L., (1984). Slow variation and characterization of domains of attraction. *Statistical Extremes and Applications*, Reidel Publication, 31-38.
- Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*. Springer, New York.
- Hill, B.I. (1975). A sample general approach to inference about the tail of a distribution. *Annals of Statistics*, **3**, 1163-1174.
- Ling, C.X., Peng, Z. and Nadarajah, S. (2007a). A location invariant moment-type estimator (I). *Theory Probability and Mathematical Statistics*, **76**, 22-30.
- Ling, C.X., Peng, Z. and Nadarajah, S. (2007b). A location invariant moment-type estimator (I). *Theory Probability and Mathematical Statistics*, **77**, 167-178.
- Pickands, J. (1975). Statistical Inference using extreme value under statistics. *Annals of Statistics*, **30**, 119-131.
- Reiss, R.D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values*, 2nd Birkhouser verlag, Basel.
- Resnick, S. (1977). Heavy tail modeling and teletraffic data. *Annals of Statistics*, **25**, 1805-1869.



A New Bivariate Exponential distribution

Z. Sharifonasabi ¹, M. H. Alamatsaz ², I. Kazemi ³

^{1,2,3}Department of Statistics, University of Isfahan, Isfahan

Abstract: In this paper, we present a new positive quadrant dependent copula and study its different properties. Then, we introduce a class of absolutely continuous bivariate distributions whose univariate margins are exponential and study its various statistical properties. To estimate the parameters, we use the method of Inference Function for Margins. Monte Carlo simulations are performed to illustrate the behavior of the estimates of the parameters. Finally, a data analysis has been performed to illustrate the flexibility of the proposed distribution.

Keywords Copula, Dependence, Bivariate distribution, Exponential distribution, Simulation.

Mathematics Subject Classification (2010): 60E, 62E.

1 Introduction

Copulas are powerful tools to model multivariate distributions and to take into account the dependence structure among random variables independent of marginal distributions. Recently, various authors have introduced different classes of new copulas to improve the performance of associated bivariate distributions, such as [Amini et al. \(2011\)](#) and [Klien and Christa \(2011\)](#). Moreover, some authors have used copula methods to construct bivariate distributions such as [Mirhosseini et al. \(2014\)](#).

As univariate exponential distributions are important in describing the lifetime of a single component, bivariate distributions with exponential marginals are used quite extensively in describing the lifetimes of two components together. Hence, bivariate exponential distributions are studied by many authors. Surveys of bivariate exponential distributions and their applications in reliability can be found in [Balakrishnan and Lai \(2009\)](#). One of the purposes for such constructions is to increase the flexibility in modeling.

The rest of this paper is organized as follows. In Section 2, we briefly review the copula concept and consider a new class of copulas and explore their properties. In Section 3, we propose a bivariate distribution whose univariate margins are exponential and study its different properties. Moreover, we provide some simulation and real data analysis for illustrating purposes.

¹Speaker's Name : Zahra Sharifonasabi zahra.sharif@sci.ac.ir

2 Copulas

A copula is a multivariate function that links univariate marginal distribution functions (df) to construct a multivariate df. For a bivariate case, according to Sklar's Theorem, if two random variables X and Y follow arbitrary marginal df's $F_X(x)$ and $F_Y(y)$, respectively, then there exists a copula, C , that combines these two marginals to give the joint df, $H_{X,Y}(x, y)$, as follows (see, [Nelsen \(2006\)](#))

$$H_{X,Y}(x, y) = C(F_X(x), F_Y(y)). \quad (2.1)$$

Conversely, if C is a copula and $F_X(x)$ and $F_Y(y)$ are df's, then the function H defined by (2.1) is a joint df with margins $F_X(x)$ and $F_Y(y)$. Furthermore, if $F_X(x)$ and $F_Y(y)$ are absolutely continuous, then C is unique. Under the assumption that the marginal distributions are continuous with probability density functions $f_X(x)$ and $f_Y(y)$, the joint probability density function (pdf) becomes $h_{X,Y}(x, y) = f_X(x)f_Y(y)c(F_X(x), F_Y(y))$, where c is the density function of C , defined as $c(u, v) = \frac{\partial^2}{\partial v \partial u} C(u, v)$. We easily can show that the bivariate function

$$C(u, v; \theta, \delta) = uv[1 + \delta(1 - u)^2(1 - v)^2][1 + \theta(1 - u)(1 - v)], \quad (2.2)$$

is a copula for $\theta \in [0, 1]$, $\delta \in [0, 1]$. The density copula is given by

$$\begin{aligned} c(u, v) &= 1 + \theta + \delta + \delta\theta - (4\delta + 2\theta + 6\theta\delta)(u + v) + (3\delta + 9\theta\delta)(u^2 + v^2) - 4\theta\delta(u^3 + v^3) \\ &- (12\delta + 54\theta\delta)(uv^2 + u^2v) + 24\theta\delta(uv^3 + u^3v) - 36\theta\delta(u^2v^3 + u^3v^2) \\ &+ (16\delta + 4\theta + 36\theta\delta)uv + (9\delta + 81\theta\delta)u^2v^2 + 16\theta\delta u^3v^3 = \sum_{m=0}^3 \sum_{n=0}^3 a_{n,m} u^n v^m, \end{aligned}$$

where $a_{n,m} = a_{n,m}(\theta, \delta)$ is constant w.r.t. u and v .

Definition 2.1. Let r.v.'s X and Y be related by a copula $C(u, v)$. Then,

1. X and Y are Positively Quadrant Dependent (PQD) (Negatively Quadrant Dependent (NQD)), if $C(u, v) \geq (\leq) uv$, $\forall (u, v) \in I^2$.
2. Y is Stochastically Increasing (SI) in X , if $C(u, v)$ is a concave function of u for all $v \in I$ (Corollary 5.2.11 in [Nelsen \(2006\)](#)).
3. Y is Left Tail Decreasing (LTD) in X , if $\frac{C(u,v)}{u}$ in u is non-increasing for all $v \in I$ (Theorem 5.2.5 in [Nelsen \(2006\)](#)).
4. X and Y are Left Corner Set Decreasing (LCS D), if C is Totally Positive of order 2 (see, e.g., Corollary 5.2.17 in [Nelsen \(2006\)](#)), i.e., for all $0 \leq u_1 \leq u_2 \leq 1$ and $0 \leq v_1 \leq v_2 \leq 1$,

$$C(u_1, v_1)C(u_2, v_2) - C(u_1, v_2)C(u_2, v_1) \geq 0.$$

5. The upper tail dependence and lower tail dependence λ_U λ_L are defined by

$$\lambda_U = 2 - \lim_{u \rightarrow 1^-} \frac{1 - C(u, u)}{1 - u}, \quad \lambda_L = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}.$$

6. The copula $C_{\underline{\theta}}$ is positively ordered with respect to $\underline{\theta}$ if for $\underline{\theta}_1 \leq \underline{\theta}_2$, we have $C_{\underline{\theta}_1} \leq C_{\underline{\theta}_2}$.

Theorem 2.2. If r.v.'s X and Y are related by the new copula defined in (2.2), then

i. X and Y are PQD, LCSD and Y is LTD in X .

ii. X and Y have no tail dependence.

iii. $C_{\theta, \delta}$ is positively ordered with respect to θ and δ .

The population version of the common nonparametric measures of association between the components of a continuous random pair (X, Y) are Kendall's tau (τ) and Spearman's rho (ρ_s) which depend only on the copula C . They are given by

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1, \quad \rho_s = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3.$$

Theorem 2.3. For Copula C in (2.2), τ and ρ_s are as follows:

$$\tau = \frac{2\theta}{9} + \frac{\delta}{18} + \frac{\delta\theta}{45} + \frac{\delta\theta^2}{450} + \frac{2\delta^2\theta}{11025}, \quad \rho_s = \frac{3\delta\theta}{100} + \frac{\delta}{12} + \frac{\theta}{3}.$$

Remark 2.4. Note that as a consequence of PQD property of the copula defined in (2.2), the range of τ and ρ_s for proposed copula defined in (2.2) are $0 \leq \tau \leq 0.302$ and $0 \leq \rho_s \leq 0.456$.

3 A bivariate exponential distribution

In this section we introduce a new bivariate distribution with marginal exponential distributions. Using (2.2) with marginal df's $F_X(x) = 1 - e^{-\lambda_1 x}$ and $G_Y(y) = 1 - e^{-\lambda_2 y}$, we obtain a bivariate exponential distribution of the form

$$H_{X,Y}(x, y) = (1 - e^{-\lambda_1 x})(1 - e^{-\lambda_2 y})(1 + \delta e^{-2\lambda_1 x - 2\lambda_2 y})(1 + \theta e^{-\lambda_1 x - \lambda_2 y}), \quad x, y \geq 0, \quad (3.1)$$

for $\lambda_1, \lambda_2 > 0, 0 \leq \theta \leq 1, 0 \leq \delta \leq 1$. The joint density function is given by

$$h_{X,Y}(x, y) = \lambda_1 \lambda_2 e^{-\lambda_1 x - \lambda_2 y} \sum_{m=0}^3 \sum_{n=0}^3 a_{n,m} (1 - e^{-\lambda_1 x})^n (1 - e^{-\lambda_2 y})^m, \quad (3.2)$$

where $a_{n,m}$ are bivariate functions of δ and θ .

Theorem 3.1. *If the random pair (X, Y) has the bivariate density defined by (3.2), then a) the joint mgf of (X, Y) , for $|t| < 4\lambda_1$ and $|s| < 4\lambda_2$, is as follows:*

$$M_{X,Y}(t, s) = \lambda_1 \lambda_2 \sum_{n=0}^3 \sum_{m=0}^3 \sum_{i=0}^n \sum_{j=0}^m \frac{a_{n,m} \binom{n}{i} \binom{m}{j} (-1)^{i+j}}{(t - (i + 1)\lambda_1)(s - (j + 1)\lambda_2)}.$$

b) the Prearon's correlation coefficient of (X, Y) is given by

$$Corr(X, Y) = \sum_{n=0}^3 \sum_{m=0}^3 \sum_{i=0}^n \sum_{j=0}^m \frac{a_{n,m} \binom{n}{i} \binom{m}{j} (-1)^{i+j}}{(i + 1)^2 (j + 1)^2} - 1.$$

c) the conditional moments are given by

$$E(Y^k | X = x) = \frac{1}{\lambda_2^k} \sum_{n=0}^3 \sum_{m=0}^3 \sum_{j=0}^m \frac{\binom{m}{j} (-1)^{j k!}}{(j + 1)^{k+1}} a_{n,m} (1 - e^{-\lambda_1 x})^n.$$

The stress-strength term comes from a reliability problem. It describes the life of a component which has a random strength X and is subject to a random stress Y . If the stress exceeds the strength $X < Y$, the component will fail while the component works whenever $X > Y$. Thus, $T = P(X < Y)$ measures of component failure and $R = P(X > Y)$ is a measure of component reliability. Many applications of the stress-strength model are related to engineering and psychology problems. The following theorem gives a form for the stress-strength model of the bivariate density in (3.2).

Theorem 3.2. *If the random pair (X, Y) has the bivariate density defined by (3.2), then*

$$P(X > Y) = \lambda_1 \sum_{m=0}^3 \sum_{n=0}^3 \sum_{i=0}^n \sum_{j=0}^{m+1} \frac{a_{n,m} \binom{n}{i} \binom{m+1}{j} (-1)^{i+j}}{(m + 1)(\lambda_1(i + 1) + \lambda_2 j)}.$$

3.1 Estimation of parameters

To estimate the parameters in a bivariate distribution, the method of Inference Function for Margins (IFM) (see, Joe (1997)) is employed. Two separate estimation procedures are involved in this method: (1) First, the parameters of the marginals are estimated from the observed values and then (2) each parametric margin is plugged into the copula likelihood which is then maximized with respect to the copula parameters. For the bivariate distribution, the log-likelihood function can be written as

$$\ln L(x, y; \lambda_1, \lambda_2, \theta, \delta) = \ln L_C(F_X(x), F_Y(y); \theta, \delta) + L_X(x; \lambda_1) + \ln L_Y(y; \lambda_2) \tag{3.3}$$

where L_C is the log-likelihood function of the density copula. Substituting the estimated parameters $\hat{\lambda}_1$ and $\hat{\lambda}_2$ into Equation (3.3), the log-likelihood function $\ln L$ is maximized to obtain the estimate of

the copula parameters, θ and δ . Hence, L_C is given by

$$\ln L_C(F_X(x), F_Y(y); \theta, \delta) = \sum_{k=1}^N \ln \left(\sum_{m=0}^3 \sum_{n=0}^3 a_{n,m} (1 - e^{-\hat{\lambda}_1 x_k})^n (1 - e^{-\hat{\lambda}_2 y_k})^m \right) \quad (3.4)$$

The maximum likelihood estimates can be obtained by maximizing (3.4) with respect to the θ and δ .

3.2 Simulation results

In this section, we obtain some simulation results to see how estimates treat for different parameter values. In order to simulate a vector (X, Y) from the bivariate distribution in (3.1), it is sufficient to simulate a vector (U, V) from copula defined in (2.2), where random variables U and V are uniform on $[0, 1]$. By using Sklar's Theorem and the Probability Integral Transform

$$U = F(X) \Leftrightarrow X = F^{[-1]}(U), \quad V = G(Y) \Leftrightarrow Y = G^{[-1]}(V),$$

where the random variables X and Y have distribution functions F and G , respectively and joint distribution H in (3.1). Hence, the simulation procedure of a copula is as follows:

- Step 1 Generate two variables U and Z independently from uniform distribution.
- Step 2 Calculate $V = C_u^{-1}(Z)$, where $C_u(v) = C(v|U = u) = \frac{\partial C(u,v)}{\partial u}$. Then, the pair (U, V) has the desired copula.

Now, we simulate data from bivariate distribution in (3.1) with two set of the parameters $(\theta, \delta, \lambda_1, \lambda_2) = (0.85, 0.5, 3, 4)$ and $(\theta, \delta, \lambda_1, \lambda_2) = (0.45, 0.2, 2, 1)$ with sample sizes 50, 150 and 200. The simulations consist of 1000 replications. In each case we estimate parameters and compute the average estimates (AE), bias and root of mean square error (RMSE). Results are given in Table 1. We can see that the performances of estimates are satisfactory. Moreover, as sample size increases RMSE decreases for all parameters. The AE of the parameters δ , λ_1 and λ_2 decreases in all cases and AE of the parameter θ increases in some cases.

3.3 Data analysis

In this section we provide the analysis of a real data set from [McGilchrist and Aisbett \(1991\)](#). The data has been obtained from an experiment, where N individuals are observed and times between recurrence of particular type of event are recorded. The first step in bivariate distribution fitting consists of identification of appropriate marginal distributions for each dependent variate. The random variables X and Y are fitted by exponential distributions. The estimated parameters by the maximum

Table 1: AE, Bias and RMSE based on 1000 simulations of the bivariate distribution in (3.1) for n=50, 150 and 200.

n		$(\theta, \delta, \lambda_1, \lambda_2) = (0.85, 0.5, 3, 4)$			$(\theta, \delta, \lambda_1, \lambda_2) = (0.45, 0.2, 2, 1)$		
		AE	Bias	RMSE	AE	Bias	RMSE
50	$\hat{\theta}$	0.6695	-0.1805	0.4398	0.4552	-0.0052	0.4359
	$\hat{\delta}$	0.4986	-0.0014	0.8446	0.4482	-0.0518	1.0632
	$\hat{\lambda}_1$	3.0053	0.0053	0.425	2.0035	0.0035	0.2833
	$\hat{\lambda}_2$	4.0133	0.0133	0.5678	1.0027	0.0027	0.1418
150	$\hat{\theta}$	0.7652	-0.0848	0.2497	0.4558	0.0058	0.2604
	$\hat{\delta}$	0.4851	-0.0149	0.4938	0.4105	0.2105	0.6929
	$\hat{\lambda}_1$	2.9984	-0.0016	0.2448	1.9989	-0.0011	0.1632
	$\hat{\lambda}_2$	3.9935	-0.0065	0.3262	0.998	-0.002	0.0815
200	$\hat{\theta}$	0.7884	-0.0616	0.2126	0.4504	0.0004	0.2288
	$\hat{\delta}$	0.4807	-0.0193	0.4282	0.3886	0.1886	0.6022
	$\hat{\lambda}_1$	2.998	-0.002	0.212	1.9987	-0.0013	0.1413
	$\hat{\lambda}_2$	3.9932	-0.0068	0.2825	0.998	-0.002	0.0706

likelihood method are $\hat{\lambda}_1 = 0.0093$ and $\hat{\lambda}_2 = 0.0086$. The Kolmogorov-Smirnov goodness-of-fit test was used to detect whether or not the proposed models could be used to represent the observed data. The statistic values for X and Y are 1.201 and 1.105 with p-value 0.093 and 0.148, respectively, that is the fitted models were acceptable. Also Figure 1 shows the comparison between the observed data and the fitted exponential distributions. Now, to estimate the parameters of copula we shall maximize (3.4) with respect to θ and δ . We use a SAS software to estimate parameters numerically. The estimated parameters by the Newton-Raphson method are $\hat{\theta} = 0.3778$ and $\hat{\delta} = 0.8532$ with standard errors 0.6009 and 1.6572, respectively. The natural question here is whether the proposed model fits this bivariate data or not. Genest et al. (2008) review and compare many proposed goodness-of-fit testings for copula models. Cramer-von Mises statistics is defined as:

$$S_n = \sum_{i=1}^n (C(u_i, v_i; \hat{\theta}, \hat{\delta}) - C_n(u_i, v_i))^2,$$

where $C(\cdot, \cdot; \hat{\theta}, \hat{\delta})$ and $C_n(\cdot, \cdot)$ are the fitted copula and empirical copula (see, Nelsen (2006)) of the data at hand, respectively, and

$$u_i = \frac{1}{n+1} \times \text{Rank of } X_i \text{ among } X_1, X_2, \dots, X_n \quad \text{and} \quad v_i = \frac{1}{n+1} \times \text{Rank of } Y_i \text{ among } Y_1, Y_2, \dots, Y_n.$$

This statistic measures how close the fitted copula is from the empirical copula of data. We use procedure in Appendix A of Genest et al. (2008). In this procedure, we have to calculate $S_1^*, S_2^*, \dots, S_{1000}^*$

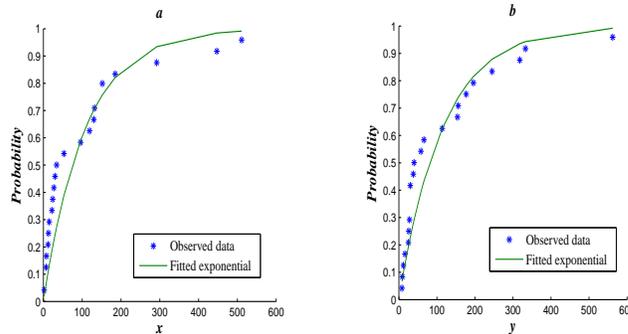


Figure 1: A comparison of the observed data and the fitted exponential distribution.

by simulated data and P-value is the proportion of these values that are larger than $S_n = 0.0596765$ which is 0.16. Thus we can conclude that the copula defined by (2.2) with the parameters $\theta = 0.3778$ and $\delta = 0.8532$ performs a good fit to this data set. Moreover, by Sklar's Theorem, an adequate bivariate distribution for this data is given by

$$H(x, y) = (1 - e^{-0.0093x})(1 - e^{-0.0086y})(1 + 0.8532e^{-0.0186x - 0.0172y}) \\ \times (1 + 0.3778e^{-0.0093x - 0.0086y}), \quad x > 0, y > 0.$$

Conclusions

We have introduced a bivariate exponential distribution whose univariate marginal distributions are exponential, discussed its measures of associations, moments and stress-strength measure. Results from the simulation of new distribution show that, for most cases the estimates are satisfactory. We analyzed a real data set and observed that the proposed model provides a good fit to the data.

References

- Amini, M., Jabbari, H. and Mohtashami Borzadaran, G. R. (2011). Aspects of dependence in generalized Farlie-Gumbel-Morgenstern distribution. *Communications in Statistics - Simulation and Computation*, **40**, 1192-1205.
- Balakrishnan, N. and Lai, C. D. (2009). *Continuous Bivariate Distribution*, Second edition. Springer, New York.

- Genest, C., Remillard, B. and Beaudion, D. (2008). Goodness-of-fit tests for copulas: A review and a power study, *Insurance: Mathematics and Economics*, **42**, 199-213.
- Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman and Hall, New York.
- Klein, I., Christa, F., (2011). Families of copulas closed under the construction of generalized linear means, *IWQW discussion paper series*, No. 04.
- McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis, *Biometrics*, **46** 461-466.
- Mirhosseini, S. M., Amini, M., Kundu, D. and Dolati, A. (2014). To appear. On a new absolutely continuous bivariate generalized exponential distribution. *Statistical Methods and Applications*, **4**, 1-26.
- Nelsen, R.B. (2006). *An introduction to copulas*. Springer, New York.



Expected Number of Real Roots of Certain Gaussian Random Trigonometric Polynomials

Soudabeh Shemehsavar¹

¹School of Mathematics, Statistics and Computer Science, University of Tehran,
Tehran, Iran

Abstract: We consider a random trigonometric polynomial where the coefficients form a sequence of Gaussian random variables. By assuming that the increments are independent, we provide the asymptotic behavior of the expected number of real roots of random trigonometric polynomial as order $\frac{2\sqrt{2n}}{\sqrt{3}}$. Also by the symmetric property assumption of coefficients, we show that the expected number of real roots is of order $\frac{2n}{\sqrt{3}}$.

Keywords Random Trigonometric Polynomials, Brownian motion, Symmetric Property.

Mathematics Subject Classification (2010): 65H42 60G99.

1 Introduction

There are two different forms of random trigonometric polynomials previously studied.

$$T_n(\theta) = \sum_{k=0}^n A_k \cos(k\theta)$$

and

$$D_n(\theta) = \sum_{k=0}^n (A_k \cos k\theta + B_k \sin k\theta), \quad (1.1)$$

For various aspects on random polynomials see Bharucha-Reid and Sambandham [Bharucha-Ried and Sambandham \(1986\)](#), which includes a comprehensive reference. Farahmand and T.Li [Farahmand and Li \(2010\)](#) obtained asymptotic behavior for the expected number of real roots of two different forms of random trigonometric polynomials $T_n(\theta)$ and $D_n(\theta)$, where the coefficients of polynomials are normally distributed random variables with different means and variances. Also They studied a case of reciprocal random polynomials for $T_n(\theta)$ and $D_n(\theta)$. We consider the classical forms of random trigonometric polynomials $D_n(\theta)$ where the coefficients A_0, A_1, \dots, A_n and B_0, B_1, \dots, B_n be a mean zero Gaussian random sequence in which the increments $\Delta_k^{(1)} = A_k - A_{k-1}$ and $\Delta_k^{(2)} = B_k - B_{k-1}$, $k = 0, 1, 2, \dots$,

¹Soudabeh Shemehsavar : shemehsavar@khayam.ut.ac.ir

are independent, $A_{-1} = 0, B_{-1} = 0$. The sequence A_0, A_1, \dots and B_0, B_1, \dots may be considered as successive Brownian points, i.e., $A_k = W_1(t_k), B_k = W_2(t_k), k = 0, 1, \dots, n$, where $t_0 < t_1 < \dots$ and $\{W_i(t_k), t \geq 0\}, i = 1, 2$, are the standard Brownian motion. In this physical interpretation, $\text{Var}(\Delta_k^{(i)})$ is the distance between successive times t_{k-1}, t_k . We note that

$$A_k = \Delta_0^{(1)} + \Delta_1^{(1)} + \dots + \Delta_k^{(1)}, \quad B_k = \Delta_0^{(2)} + \Delta_1^{(2)} + \dots + \Delta_k^{(2)} \quad k = 0, 1, \dots, n$$

, where $\Delta_k^{(i)} \sim N(0, \sigma_i^2), k = 0, 1, \dots, n, i = 1, 2$, and $\Delta_k^{(i)}$ are independent. Thus

$$D_n(\theta) = \sum_{k=0}^n \left[\left(\sum_{j=k}^n \cos j\theta \right) \Delta_k^{(1)} + \left(\sum_{j=k}^n \sin j\theta \right) \Delta_k^{(2)} \right] = \sum_{k=0}^n \left(a_{k1}(\theta) \Delta_k^{(1)} + b_{k1}(\theta) \Delta_k^{(2)} \right)$$

and

$$D'_n(\theta) = \sum_{k=0}^n \left[\left(- \sum_{j=k}^n j \sin j\theta \right) \Delta_k^{(1)} + \left(\sum_{j=k}^n j \cos j\theta \right) \Delta_k^{(2)} \right] = \sum_{k=0}^n \left(c_{k1}(\theta) \Delta_k^{(1)} + d_{k1}(\theta) \Delta_k^{(2)} \right)$$

Where

$$\begin{aligned} a_{k1}(\theta) &= \sum_{j=k}^n \cos j\theta = \frac{\sin(2n+1)\frac{\theta}{2} - \sin(2k-1)\frac{\theta}{2}}{2 \sin(\frac{\theta}{2})}, \quad b_{k1}(\theta) = \sum_{j=k}^n \sin j\theta = \frac{\cos(2k-1)\frac{\theta}{2} - \cos(2n+1)\frac{\theta}{2}}{2 \sin(\frac{\theta}{2})}, \\ c_{k1}(\theta) &= - \sum_{j=k}^n j \sin j\theta = \left(\frac{\sin(2n+1)\frac{\theta}{2} - \sin(2k-1)\frac{\theta}{2}}{2 \sin(\frac{\theta}{2})} \right)', \\ d_{k1}(\theta) &= \sum_{j=k}^n j \cos j\theta = \left(\frac{\cos(2k-1)\frac{\theta}{2} - \cos(2n+1)\frac{\theta}{2}}{2 \sin(\frac{\theta}{2})} \right)' \end{aligned} \tag{1.2}$$

Now given $D_n(\theta)$ in (1.1) with a symmetric property of coefficients, i.e., $A_k \equiv A_{n-k}$ and $B_k \equiv B_{n-k}$ for $k = 0, 1, \dots, n$, we can write $Q_n(\theta)$ for odd n 's as follows:

$$Q_n(\theta) = \sum_{k=0}^{\frac{n-1}{2}} [A_k(\cos k\theta + \cos(n-k)\theta) + B_k(\sin k\theta + \sin(n-k)\theta)] \tag{1.3}$$

The polynomials will have one additional term for even n 's and we will not discuss this case here.

$$Q_n(\theta) = \sum_{k=0}^{\frac{n-1}{2}} [A_k(\cos k\theta + \cos(n-k)\theta) + B_k(\sin k\theta + \sin(n-k)\theta)]$$

$$\begin{aligned}
&= \sum_{k=0}^{\frac{n-1}{2}} \left[\left(\sum_{j=k}^{\frac{n-1}{2}} (\cos j\theta + \cos(n-j)\theta) \right) \Delta_k^{(1)} + \left(\sum_{j=k}^{\frac{n-1}{2}} (\sin j\theta + \sin(n-j)\theta) \right) \Delta_k^{(2)} \right] \\
&= \sum_{k=0}^{\frac{n-1}{2}} \left(a_{k2}(\theta) \Delta_k^{(1)} + b_{k2}(\theta) \Delta_k^{(2)} \right), \\
Q'_n(\theta) &= \sum_{k=0}^{\frac{n-1}{2}} \left[\left(- \sum_{j=k}^{\frac{n-1}{2}} (j \sin j\theta + (n-j) \sin(n-j)\theta) \right) \Delta_k^{(1)} \right. \\
&\quad \left. + \left(\sum_{j=k}^{\frac{n-1}{2}} (j \cos j\theta + (n-j) \cos(n-j)\theta) \right) \Delta_k^{(2)} \right] = \sum_{k=0}^{\frac{n-1}{2}} \left(c_{k2}(\theta) \Delta_k^{(1)} + d_{k2}(\theta) \Delta_k^{(2)} \right)
\end{aligned}$$

where by using this results

$$\begin{aligned}
a_{k2}(\theta) &= \sum_{j=k}^{\frac{n-1}{2}} (\cos j\theta + \cos(n-j)\theta) = \frac{\sin(2n-2k+1)\frac{\theta}{2} - \sin(2k-1)\frac{\theta}{2}}{2 \sin \frac{\theta}{2}}, \\
b_{k2}(\theta) &= \sum_{j=k}^{\frac{n-1}{2}} (\sin j\theta + \sin(n-j)\theta) = \frac{\cos(2k-1)\frac{\theta}{2} - \cos(2n-2k+1)\frac{\theta}{2}}{2 \sin \frac{\theta}{2}}, \\
c_{k2}(\theta) &= - \sum_{j=k}^{\frac{n-1}{2}} (j \sin j\theta + (n-j) \sin(n-j)\theta) = \left(\frac{\sin(2n-2k+1)\frac{\theta}{2} - \sin(2k-1)\frac{\theta}{2}}{2 \sin \frac{\theta}{2}} \right)', \\
d_{k2}(\theta) &= \sum_{j=k}^{\frac{n-1}{2}} (j \cos j\theta + (n-j) \cos(n-j)\theta) = \left(\frac{\cos(2k-1)\frac{\theta}{2} - \cos(2n-2k+1)\frac{\theta}{2}}{2 \sin \frac{\theta}{2}} \right)', \quad (1.4)
\end{aligned}$$

2 Kac-Rice Formula

Let $N(0, 2\pi)$ be denotes the number of real roots of the random trigonometric polynomials in the interval $(0, 2\pi)$ and $E(N(0, 2\pi))$ be its expected value. To deal with the asymptotic behavior of the expected number of real roots of $D_n(\theta) = 0$ and $Q_n(\theta) = 0$, we refer to Kac-Rice formula [Farahmand \(1998\)](#), which is defined as

$$E(N(0, 2\pi)) = \int_0^{2\pi} \frac{\Delta}{\pi A^2} d\theta \quad (2.1)$$

where $\Delta^2 = A^2B^2 - C^2$. For $D_n(\theta)$ given in (1.1) we have

$$A_D^2 = \text{Var}(D_n(\theta)) = \sum_{k=0}^n (a_{k1}^2(\theta)\sigma_1^2 + b_{k1}^2(\theta)\sigma_2^2), \quad B_D^2 = \text{Var}(D'_n(\theta)) = \sum_{k=0}^n (c_{k1}^2(\theta)\sigma_1^2 + d_{k1}^2(\theta)\sigma_2^2),$$

$$C_D = \text{Cov}(D_n(\theta), D'_n(\theta)) = \sum_{k=0}^n (a_{k1}(\theta)c_{k1}(\theta)\sigma_1^2 + b_{k1}(\theta)d_{k1}(\theta)\sigma_2^2), \tag{2.2}$$

where $a_{k1}(\theta)$, $b_{k1}(\theta)$, $c_{k1}(\theta)$ and $d_{k1}(\theta)$ are defined in (1.2). For $Q_n(\theta)$ given in (1.3) we have

$$A_Q^2 = \text{Var}(Q_n(\theta)) = \sum_{k=0}^{\frac{n-1}{2}} (a_{k2}^2(\theta)\sigma_1^2 + b_{k2}^2(\theta)\sigma_2^2), \quad B_Q^2 = \text{Var}(Q'_n(\theta)) = \sum_{k=0}^{\frac{n-1}{2}} (c_{k2}^2(\theta)\sigma_1^2 + d_{k2}^2(\theta)\sigma_2^2),$$

$$C_Q = \text{Cov}(Q_n(\theta), Q'_n(\theta)) = \sum_{k=0}^{\frac{n-1}{2}} (a_{k2}(\theta)c_{k2}(\theta)\sigma_1^2 + b_{k2}(\theta)d_{k2}(\theta)\sigma_2^2), \tag{2.3}$$

where $a_{k2}(\theta)$, $b_{k2}(\theta)$, $c_{k2}(\theta)$ and $d_{k2}(\theta)$ are defined in (1.4).

As in algebraic case the above identities are not well behaved around $0, \pi$ and 2π . Therefore we first consider the intervals $(\varepsilon, \pi - \varepsilon), (\pi + \varepsilon, 2\pi - \varepsilon)$, where ε is any positive constant, smaller than π and arbitrary at this point to be chosen later. It should be positive and small enough to facilitate handling the roots in the intervals $(\varepsilon, \pi - \varepsilon), (\pi + \varepsilon, 2\pi - \varepsilon)$ and for roots inside this two intervals, we use (2.1). For the real roots lying in the intervals $(0, \varepsilon), (\pi - \varepsilon, \pi + \varepsilon)$ and $(2\pi - \varepsilon, 2\pi)$, which it so happens, are negligible, we will use a different method based on the Jensen's theorem.

We now define some functions to make the estimations, define $S(\theta) = \sin(2n + 1)\theta / \sin \theta$, see from [6,page 74] which is continuous at $\theta = j\pi$ and will occur frequently in follows. Since for $\theta \in (\varepsilon, \pi - \varepsilon)$ and $\theta \in (\pi + \varepsilon, 2\pi - \varepsilon)$, we have $|S(\theta)| < 1 / \sin \varepsilon$. Hence, we can obtain

$$S(\theta) = O\left(\frac{1}{\varepsilon}\right), \quad S'(\theta) = O\left(\frac{n}{\varepsilon}\right), \quad S''(\theta) = O\left(\frac{n^2}{\varepsilon}\right)$$

We can show

$$\sum_{k=0}^n \cos k\theta = \frac{\sin(2n + 1)\frac{\theta}{2}}{2 \sin \frac{\theta}{2}} + \frac{1}{2} = \frac{S(\frac{\theta}{2}) + 1}{2} = O\left(\frac{1}{\varepsilon}\right) \tag{2.4}$$

and

$$\sum_{k=0}^n k \sin k\theta = -\frac{S'(\frac{\theta}{2})}{4} = O\left(\frac{n}{\varepsilon}\right), \quad \sum_{k=0}^n k^2 \cos k\theta = -\frac{S''(\frac{\theta}{2})}{8} = O\left(\frac{n^2}{\varepsilon}\right) \tag{2.5}$$

In similar way, we define $P(\theta) = \cos \theta - \frac{\cos(2n+1)\theta}{2 \sin \theta}$, we also have $|P(\theta)| < 1 / \sin \varepsilon$. Hence, we can obtain

$$P(\theta) = O\left(\frac{1}{\varepsilon}\right), \quad P'(\theta) = O\left(\frac{n}{\varepsilon}\right), \quad P''(\theta) = O\left(\frac{n^2}{\varepsilon}\right)$$

We can show

$$\sum_{k=0}^n \sin k\theta = \frac{\cos \frac{\theta}{2} - \cos(2n+1)\frac{\theta}{2}}{2 \sin \frac{\theta}{2}} = P\left(\frac{\theta}{2}\right) = O\left(\frac{1}{\varepsilon}\right) \tag{2.6}$$

and

$$\sum_{k=0}^n k \cos k\theta = \frac{P'\left(\frac{\theta}{2}\right)}{4} = O\left(\frac{n}{\varepsilon}\right), \quad \sum_{k=0}^n k^2 \sin k\theta = -\frac{P''\left(\frac{\theta}{2}\right)}{8} = O\left(\frac{n^2}{\varepsilon}\right) \tag{2.7}$$

Now, using the above identities, we are able to evaluate the characteristics required in using the Kac-Rice formula in (2.1).

3 Asymptotic Behavior of $E(N(0, 2\pi))$

This section includes two subsection. We evaluate the asymptotic behavior of the expected number of real roots of $D_n(\theta) = 0$ in the intervals $(\varepsilon, \pi - \varepsilon)$, $(\pi + \varepsilon, 2\pi - \varepsilon)$ in subsection 3.1 and in the intervals $(0, \varepsilon)$, $(\pi - \varepsilon, \pi + \varepsilon)$, $(2\pi - \varepsilon, 2\pi)$ in subsection 3.2.

3.1 Results On the Intervals $(\varepsilon, \pi - \varepsilon)$, $(\pi + \varepsilon, 2\pi - \varepsilon)$

In this part, we obtain our results by applying the Kac-Rice formula.

Theorem 3.1. *Let $D_n(\theta)$ be the random trigonometric polynomial given in (1.1) for which $A_k = \Delta_0^{(1)} + \Delta_1^{(1)} + \dots + \Delta_k^{(1)}$, $B_k = \Delta_0^{(2)} + \Delta_1^{(2)} + \dots + \Delta_k^{(2)}$, $k = 0, 1, \dots, n$, where $\Delta_k^{(i)}$, $k = 0, 1, \dots, n$, $i = 1, 2$ are standard normal i.i.d random variables independent. We prove that for all sufficiently large n , the expected number of real roots of the equation $D_n(\theta) = 0$, satisfies*

$$EN(\varepsilon, \pi - \varepsilon) = EN(\pi + \varepsilon, 2\pi - \varepsilon) \simeq \frac{\sqrt{2}n}{\sqrt{3}}$$

Proof. In order to use the Kac-Rice formula, we first evaluate asymptotic value for each variable needed by using the error terms obtained in (2.4)-(2.7). Since $E(A_k) = 0$ and $E(B_k) = 0$ we have

$$E(D_n(\theta)) = 0, \quad E(D'_n(\theta)) = 0, \tag{3.1}$$

Now using (2.4)-(2.7) and (1.2) and using some trigonometric identities, we obtain the variance of $D_n(\theta)$ and $D'_n(\theta)$, as

$$A_D^2 = \text{Var}(D_n(\theta)) = \sum_{k=0}^n (a_{k1}^2(\theta) + b_{k1}^2(\theta)) = \frac{n}{2 \sin^2 \frac{\theta}{2}} + O\left(\frac{1}{\varepsilon}\right), \tag{3.2}$$

$$B_D^2 = \text{Var}(D'_n(\theta)) = \sum_{k=0}^n (c_{k1}^2(\theta) + d_{k1}^2(\theta)) = \frac{n^3}{3 \sin^2 \frac{\theta}{2}} + O\left(\frac{n^2}{\varepsilon}\right), \tag{3.3}$$

$$C_D = \text{Cov}(D_n(\theta), D'_n(\theta)) = \sum_{k=0}^n (a_{k1}(\theta)c_{k1}(\theta) + b_{k1}(\theta)d_{k1}(\theta)) = O\left(\frac{n}{\varepsilon}\right) \tag{3.4}$$

Then, finally from (3.2)-(3.4), we can obtain

$$\Delta^2 = A_D^2 B_D^2 - C_D^2 = \frac{n^4}{6 \sin^4 \frac{\theta}{2}} + O\left(\frac{n^3}{\varepsilon}\right), \tag{3.5}$$

The results of (3.2) and (3.5) into the Kac-Rice formula (2.1), we have

$$E(N(\varepsilon, \pi - \varepsilon)) = E(N(\pi + \varepsilon, 2\pi - \varepsilon)) \sim \frac{\sqrt{2}n}{\sqrt{3}}$$

The theorem is proved. □

Theorem 3.2. *Let $Q_n(\theta)$ be the random trigonometric polynomial given in (1.3) where $A_k = \Delta_0^{(1)} + \Delta_1^{(1)} + \dots + \Delta_k^{(1)}$, $B_k = \Delta_0^{(2)} + \Delta_1^{(2)} + \dots + \Delta_k^{(2)}$, $k = 0, 1, \dots, \frac{n-1}{2}$, where $\Delta_k^{(i)}$, $k = 0, 1, \dots, \frac{n-1}{2}$, $i = 1, 2$, are standard normal i.i.d random variables. We prove that for all sufficiently large n , the expected number of real roots of the equation $Q_n(\theta) = 0$, satisfies*

$$EN(\varepsilon, \pi - \varepsilon) = EN(\pi + \varepsilon, 2\pi - \varepsilon) \simeq \frac{n}{\sqrt{3}}$$

Proof. We obtain our results by applying the Kac-Rice formula. Since $E(A_k) = 0$ and $E(B_k) = 0$ we have $E(Q_n(\theta)) = 0$, $E(Q'_n(\theta)) = 0$. Now from (1.4) and (2.4)-(2.7) and making some trigonometric identities, we obtain

$$A_Q^2 = \text{Var}(Q_n(\theta)) = \sum_{k=0}^{\frac{n-1}{2}} (a_{k2}^2(\theta) + b_{k2}^2(\theta)) = \frac{n}{4 \sin^2 \frac{\theta}{2}} + O\left(\frac{1}{\varepsilon}\right), \tag{3.11}$$

$$B_Q^2 = \text{Var}(Q'_n(\theta)) = \sum_{k=0}^{\frac{n-1}{2}} (c_{k2}^2(\theta) + d_{k2}^2(\theta)) = \frac{n^3}{12 \sin^2 \frac{\theta}{2}} + O\left(\frac{n^2}{\varepsilon}\right), \tag{3.12}$$

and

$$C_Q^2 = \text{Cov}(Q_n(\theta), Q'_n(\theta)) = \sum_{k=0}^{\frac{n-1}{2}} (a_{k2}(\theta)c_{k2}(\theta) + b_{k2}(\theta)d_{k2}(\theta)) = O\left(\frac{n}{\varepsilon}\right), \tag{3.13}$$

Then, finally from (3.11)-(3.13) we can get

$$\Delta^2 = A_Q^2 B_Q^2 - C_Q^2 = \frac{n^4}{48 \sin^4 \frac{\theta}{2}} + O\left(\frac{n^3}{\varepsilon}\right), \tag{3.14}$$

The results of (3.11) and (3.14) into the Kac-Rice formula (2.1), we can obtain

$$E(N(\varepsilon, \pi - \varepsilon)) = E(N(\pi + \varepsilon, 2\pi - \varepsilon)) \sim \frac{n}{\sqrt{3}}$$

□

3.2 Results On the Intervals $(0, \varepsilon), (\pi - \varepsilon, \pi + \varepsilon), (2\pi - \varepsilon, 2\pi)$

In this subsection, we are going to show the expected number of real roots in the intervals $(0, \varepsilon), (\pi - \varepsilon, \pi + \varepsilon), (2\pi - \varepsilon, 2\pi)$ is negligible. The period of $D_n(\theta)$ is 2π , and so the number of real roots in the interval $(0, \varepsilon)$ and $(2\pi - \varepsilon, 2\pi)$ is the same as the number in $(-\varepsilon, \varepsilon)$, the interval $(\pi - \varepsilon, \pi + \varepsilon)$ can be treated in the same way to give the same result. Here we deal only with $D_n(\theta)$, since the same method is applicable for the random trigonometric polynomial, $Q_n(\theta)$, and the results of $D_n(\theta)$ remain the same for $Q_n(\theta)$. We consider the function of the complex variable z ,

$$D_n(z, \omega) = \sum_{k=0}^n (A_k(\omega) \cos kz + B_k(\omega) \sin kz)$$

We seek an upper bound to the number of real roots in the segment of the real axis joining the points $\pm\varepsilon$, and this certainly does not exceed the number of real roots in the circle $|z| < \varepsilon$.

let $N(r) \equiv N(r, \omega)$ denote the number of real roots of $D_n(z, \omega) = 0$ in $|z| < \varepsilon$. We will modify the method based on the Jensen's theorem [Rudin \(1974\)](#), By Jensen's theorem,

$$\int_{\varepsilon}^{2\varepsilon} r^{-1} N(r) dr \leq \int_0^{2\varepsilon} r^{-1} N(r) dr = \frac{1}{2\pi} \int_0^{2\pi} \log \left| \frac{D_n(2\varepsilon e^{i\theta}, \omega)}{D_n(0)} \right| d\theta$$

for which we have

$$N(\varepsilon) \log 2 \leq \frac{1}{2\pi} \int_0^{2\pi} \log \left| \frac{D_n(2\varepsilon e^{i\theta}, \omega)}{D_n(0)} \right| d\theta \tag{4.1}$$

Since the distribution function of $D_n(0, \omega) = \sum_{k=0}^n \sum_{j=k}^n \Delta_k^1(\omega)$ is $G(x) \sim N\left(0, \frac{(2n^3 + 9n^2 + 13n + 6)}{6}\right)$ We can see that for any positive v

$$\begin{aligned} P(-e^{-v} \leq D_n(0, \omega) \leq e^{-v}) &= \sqrt{\frac{3}{\pi(2n^3 + 9n^2 + 13n + 6)}} \int_{-e^{-v}}^{e^{-v}} \exp\left\{-\frac{3t^2}{2n^3 + 9n^2 + 13n + 6}\right\} dt \\ &< \frac{2\sqrt{3}e^{-v}}{\sqrt{\pi(2n^3 + 9n^2 + 13n + 6)}} \end{aligned} \tag{4.2}$$

Also we have

$$|D_n(2\varepsilon e^{i\theta})| = \left| \sum_{k=0}^n \left(\sum_{j=k}^n \cos(2j\varepsilon e^{i\theta}) \right) \Delta_k^1 + \sum_{k=0}^n \left(\sum_{j=k}^n \sin(2j\varepsilon e^{i\theta}) \right) \Delta_k^2 \right| \leq 2M(n+1)(n+2)e^{2n\varepsilon} \tag{4.3}$$

Where $M = \text{Max}_k(\text{max}|\Delta_k^1|, \text{max}|\Delta_k^2|)$. The distribution function of $|\Delta_k^1|$ and $|\Delta_k^2|$ is

$$F(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

For any positive v and all sufficiently large n , the probability $M > ne^v$ is

$$P(M > ne^v) \leq nP(|\Delta_1^1| > ne^v) = n \frac{1}{\sqrt{2\pi}} \int_{ne^v}^\infty e^{-\frac{t^2}{2}} dt \simeq \sqrt{\frac{2}{\pi}} \exp \left\{ -v - \frac{(ne^v)^2}{2} \right\} \tag{4.4}$$

Therefore from (4.3) and (4.4), except for sample functions in an ω -set of measure not exceeding $\sqrt{\frac{2}{\pi}} \exp \left\{ -v - \frac{(ne^v)^2}{2} \right\}$

$$|D_n(2\varepsilon e^{i\theta})| < 3n(n+1)(n+2)e^{2n\varepsilon+v} \tag{4.5}$$

Hence from (4.2), (4.5) and since we obtain

$$\left| \frac{D_n(2\varepsilon e^{i\theta}, \omega)}{D_n(0, \omega)} \right| \leq 3n(n+1)(n+2)e^{2n\varepsilon+2v} \tag{4.6}$$

Except for sample function in an ω -set of measure not exceeding

$$\frac{2\sqrt{3}e^{-v}}{\sqrt{\pi(2n^3 + 9n^2 + 13n + 6)}} + \sqrt{\frac{2}{\pi}} \exp \left\{ -v - \frac{(ne^v)^2}{2} \right\}$$

Therefore from (4.1) and (4.6) we can show that outside the exceptional set

$$N(\varepsilon) \leq \frac{\log 3 + \log n \log(n+1) + \log(n+2) + 2n\varepsilon + 2v}{\log 2} \tag{4.7}$$

$\varepsilon = n^{-1/4}$ it follows from (4.7) and for any sufficiently large n that

$$P(N(\varepsilon) > 3n\varepsilon + 2v) \leq \frac{2\sqrt{3}e^{-v}}{\sqrt{\pi(2n^3 + 9n^2 + 13n + 6)}} + \sqrt{\frac{2}{\pi}} \exp \left\{ -v - \frac{(ne^v)^2}{2} \right\} \tag{4.8}$$

Let $n' = [3n^{3/4}]$ be the greatest integer less than equal to $3n^{3/4}$, then from (4.8) and for n large enough we obtain

$$\begin{aligned} EN(\varepsilon) &= \sum_{j>0} P(N(\varepsilon) \geq j) = \sum_{1 \leq j \leq n'} P(N(\varepsilon) > j) + \sum_{j \geq 1} P(N(\varepsilon) \geq n' + j) \\ &\leq n' + \sqrt{\frac{12}{\pi(2n^3 + 9n^2 + 13n + 6)}} \sum_{j \geq 1} e^{-j/2} + \sqrt{\frac{2}{\pi}} \sum_{j \geq 1} \exp \left\{ -\frac{j}{2} - \frac{(ne^j)^2}{2} \right\} = O(n^{3/4}) \end{aligned} \tag{4.8}$$

References

- Bharucha-Ried, A. T. and Sambandham, M. (1986) , *Random Polynomials*, Academic Press, N.Y.
- Farahmand, K. (1998). *Topics in Random Polynomials*, Addison Wesley Longman, London.
- Farahmand, K. and Li,T. (2010), Random Trigonometric Polynomials with Nonidentically Distributed Coefficients,*International Journal of Stochastic Analysis*, Volume 2010, Article ID 931565, 10 pages.
- Rudin, W. (1974) . *Real and Complex Analysis*. McGraw-Hill.



Discrete Exponential INAR(1) model

M. Shirozhan¹, M. Mohammadpour²

²Department of Statistics, University of Mazandaran, Babolsar, Iran.

Abstract: In this paper, we introduce a new stationary first order integer-valued autoregressive process with discrete exponential marginal distributions. Several statistical properties of the process are established, including the autocorrelation function, spectral density function, joint probability generating function, multi-step ahead conditional expectation and variance, the one-step transition probabilities. Estimations of the model parameters are obtained by the conditional least squares, non parametric method and maximum likelihood estimation methods. The performance of the estimates of methods is checked by a small Monte Carlo simulation.

Keywords Binomial thinning, Mixture distribution, Exponential distribution, Estimation.

Mathematics Subject Classification (2010): 62m10 62M99.

1 Introduction

Integer-valued time series can model data with correlated counts and various kinds of discrete valued time series models are often encountered in practice. e.g., counts of accidents, number of transmitted messages, the number of days with storm, the number of road accidents, the number of bases of DNA sequences and so on. The most common integer-valued time series models are constructed via the binomial thinning operator $\alpha \circ$, that was first introduced by [Steutel and van Harn \(1979\)](#) in the form

$$\alpha \circ X = \sum_{i=1}^X Y_i, \quad \alpha \in (0, 1) \quad (1.1)$$

where X is a non-negative integer-valued random variable and $\{Y_i\}$ is a sequence of independent identically distributed random variables with Bernoulli(α) distribution and is independent of X . Modeling of INAR(1) time series was first introduced by [Mckenzie \(1985\)](#), and [Al-Osh and Alzaid \(1987\)](#) based on the operator $\alpha \circ$. Other models of INAR time series have been introduced based on negative binomial thinning operator generated by independent geometric counting series variables ([Aly and Bouzar \(1994\)](#), [Risti c et al. \(2009\)](#), [Risti c et al. \(2012\)](#)). Recently, [Schweer & Weib \(2014\)](#) introduced a general family of INAR(1) models with compound Poisson innovations. In this paper, the discrete

¹M. Shirozhan : ms.shirozhan@gmail.com

exponential marginal distribution is used directly to model the INAR(1) time series $\{X_t\}$ with the operator $\alpha \circ$ in (1.1) and after that distribution of the innovations of the INAR(1) is determined.

Definition 1.1. A random variable X is said to have a discrete exponential (DEXP) distribution with parameter λ , ($\lambda > 0$) denoted as $DEXP(\lambda)$, if its probability mass function (pmf) is defined as

$$p(X = x) = e^{-\lambda x} (1 - e^{-\lambda}), \quad x = 0, 1, 2, \dots$$

Proof. Consider $Y \sim exp(\lambda)$, then $X = [Y]$ have discrete exponential distribution

$$p(X = x) = p([Y] = x) = \int_x^{x+1} f_Y(y) dy = e^{-\lambda x} (1 - e^{-\lambda})$$

□

Also, the probability generating function (pgf) of DEXP distribution is given by $\varphi_X(s) = \frac{e^\lambda - 1}{e^\lambda - s}$. The rest of the paper is organized as follows. In Section 2, we construct the discrete exponential first-order integer-valued autoregressive (DEXPINAR(1)) model and obtain the pmf of innovation term of the model. In Section 3, we investigate many properties of this model, such as multi-step ahead conditional expectation variance and autocorrelation function, spectral density function, and partial autocorrelation function. In Section 4, estimations of the model parameters are obtained by the conditional least squares non parametric and maximum likelihood estimation methods. The performance of the estimates of methods is checked by Monte Carlo simulation.

2 The DEXPINAR(1) model

Consider the INAR(1) process $\{X_t\}$ as

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \quad t \geq 1 \tag{2.1}$$

where $\{X_t\}$ is stationary non-negative integer valued process following the $DEXP(\lambda)$ marginal distribution and $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables independent of $\{Y_i\}$ and $\{X_{t-l}\}$ also ε_t are independent for all $l \geq 1$. We shall refer to this model as DEXPINAR(1). From equation (2.1) and the above assumptions, the pgf of the innovation $\{\varepsilon_t\}$ is given by

$$\varphi_\varepsilon(s) = \frac{\varphi_x(s)}{\varphi_x(1 - \alpha + \alpha s)} = \frac{e^\lambda - 1 + \alpha - \alpha s}{e^\lambda - s} = \alpha + (1 - \alpha) \frac{e^\lambda - 1}{e^\lambda - s}$$

Clearly $\varphi_\varepsilon(1) = 1$, that is $\varphi_\varepsilon(s)$ is a proper pgf.

Theorem 2.1. Consider the DEXPINAR(1) process defined by (2.1), the innovation sequence $\{\varepsilon_t\}$ possesses the distribution

$$f_\varepsilon(x) = \alpha h(x) + (1 - \alpha) e^{-\lambda x} (1 - e^{-\lambda}) \quad (2.2)$$

where $h(x)$ is the degenerate distribution function at zero.

So, the innovation $\{\varepsilon_t\}$ has a mixture distribution of a degenerate distribution at 0 and discrete exponential(λ) with mixing portions α and $1 - \alpha$. Its mean and variance are

$$\mu_\varepsilon = \frac{1 - \alpha}{e^\lambda - 1}, \quad \sigma_\varepsilon^2 = \frac{(1 - \alpha)(e^\lambda + \alpha)}{(e^\lambda - 1)^2}$$

Another representation of the DEXPINAR(1) model is

$$X_t = \begin{cases} \alpha \circ X_{t-1} & w.p.\alpha \\ \alpha \circ X_{t-1} + g_t & w.p.1 - \alpha \end{cases}$$

where $\{g_t\}$ has discrete exponential (λ) distribution. The mean and the variance of X are

$$\mu_X = \frac{1}{e^\lambda - 1}, \quad \sigma_X^2 = \frac{e^\lambda}{(e^\lambda - 1)^2}$$

and hence the DEXP(λ) is over dispersed. If $X \sim DEXP(\lambda)$, then we have

$$p(X > t + s | X > S) = \frac{p(X > t + s)}{p(X > S)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda(s)}} = e^{-\lambda(t)} = p(x > t)$$

this implies that the distribution does not depend on s . This property is called the memoryless property of the discrete exponential distribution. That means I don't need to remember when I started the process. In discrete distribution only Geometric and Discrete Exponential distribution have memoryless properties.

The coefficient of variation of Discrete Exponential(λ) is $C.V_{DEXP} = e^{\frac{\lambda}{2}}$, as the coefficient of variation of the Geometric(p) is $C.V_{GEO} = \sqrt{1 - p}$. Then we conclude that $C.V_{DEXP} > C.V_{GEO}$.

3 Properties of the process

In this section we investigate some statistical and conditional properties of the DEXPINAR(1) process, involving the one-step transition probabilities, multi step ahead conditional mean and variance, autocorrelations, spectral density.

The DEXPINAR(1) process is a stationary discrete time Markov chain with the one-step transition probabilities of going from state i to state j

$$P_{ij} = p(X_t = j | X_{t-1} = i) = \sum_{k=0}^{\min(i, j-1)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} p(\varepsilon_t = j-k) \quad (3.1)$$

where $p(\varepsilon_t = j-k)$ is the pmf of $\{\varepsilon_t\}$ defined by (2.2). Conditional expectation is one of the most common techniques for forecasting time series data. Using the definition of the DEXPINAR(1) process, the $(k+1)$ -step ahead conditional mean obtained as

$$E(X_{t+k} | X_t = x) = \alpha^k x + \frac{1-\alpha^k}{1-\alpha} \mu_\varepsilon$$

and if $k \rightarrow \infty$ then $E(X_{t+k} | X_t = x) \rightarrow \frac{\mu_\varepsilon}{1-\alpha} = \frac{1}{e^\lambda - 1}$, which is the unconditional mean of the process.

The $(k+1)$ -step ahead conditional variance is

$$\text{Var}(X_{t+k} | X_t = x) = \alpha^k (1-\alpha^k) x + \sigma_\varepsilon^2 \frac{1-\alpha^{2k}}{1-\alpha^2} + \left(\frac{1-\alpha^k}{1-\alpha} - \frac{1-\alpha^{2k}}{1-\alpha^2} \right) \mu_\varepsilon$$

As $k \rightarrow \infty$, we find that

$$\text{Var}(X_{t+k} | X_t = x) \rightarrow \frac{\sigma_\varepsilon^2}{1-\alpha^2} + \left(\frac{1}{1-\alpha} - \frac{1}{1-\alpha^2} \right) \mu_\varepsilon = \frac{e^\lambda}{(e^\lambda - 1)^2} = \sigma_X^2$$

which is the unconditional variance of the process.

The auto covariance function of the process $\{X_t\}$ obtained as

$$\begin{aligned} \gamma(k) &= \text{Cov}(X_t, X_{t-k}) = \text{Cov}(\alpha \circ X_{t-1}, X_{t-k}) + \text{Cov}(\varepsilon_t, X_{t-k}) \\ &= \alpha \text{Cov}(X_{t-1}, X_{t-k}) = \dots = \alpha^k \gamma(0) \end{aligned}$$

Then the autocorrelation function is $\rho_k = \text{Corr}(X_t, X_{t-k}) = \alpha^k$.

By applying the auto covariance function, the spectral density function of the process obtains as

$$f_{xx}(\omega) = \frac{e^\lambda (1-\alpha^2)}{2\pi (e^\lambda - 1)^2 (1 + \alpha^2 - 2\alpha \cos \omega)}, \quad \omega \in (-\pi, \pi]$$

Moreover, the joint pgf of the process is obtained as follows

$$\begin{aligned} \varphi_{x_t, x_{t-1}}(s_1, s_2) &= \varphi_{\alpha \circ x_{t-1} + \varepsilon_t, x_{t-1}}(s_1, s_2) = \varphi_{x_t}(s_2(1-\alpha + \alpha s_1)) \varphi_{\varepsilon_t}(s_1) \\ &= \left(\frac{e^\lambda - 1 + \alpha - \alpha s_1}{e^\lambda - s_1} \right) \left(\frac{e^\lambda - 1}{e^\lambda - s_2(1-\alpha + \alpha s_1)} \right) \end{aligned}$$

Further, the partial autocorrelation function at lag $h > 1$ is

$$\begin{aligned}\alpha(h) &= \text{Corr}(X_{h+1} - E(X_{h+1}|X_2, \dots, X_h), X_1) \\ &= \text{Corr}\left(X_{h+1} - \alpha X_h - \frac{(1-\alpha)}{e^\lambda - 1}, X_1\right) = \alpha^h - \alpha(\alpha^{h-1}) = 0\end{aligned}$$

The one step ahead conditional probability generating function of the process is as follows

$$\begin{aligned}\varphi_{x_{t+1}|x_t=x}(s) &= \varphi_{\alpha \circ x_t + \varepsilon_{t+1}|x_t=x}(s) \\ &= E\left(s^{\alpha \circ x_t | x_t=x} s^{\varepsilon_{t+1}}\right) = (1 - \alpha + \alpha s) \frac{e^\lambda - 1 + \alpha - \alpha s}{e^\lambda - s}\end{aligned}$$

and then

$$\begin{aligned}\varphi_{x_{t+k}|x_t=x}(s) &= \varphi_{\alpha^k \circ x_t + \sum_{i=0}^{k-1} \alpha^i \circ \varepsilon_{t+k-i}|x_t=x}(s) \\ &= (1 - \alpha^k + \alpha^k s) \prod_{i=0}^{k-1} \varphi_\varepsilon(1 - \alpha^i + \alpha^i s)\end{aligned}$$

4 Estimation and simulation

Various features of the model depend on its parameters, so estimation of the model parameters is an essential topic. In this section we describe the estimation of the unknown parameters of the DEXPINAR(1) process and are compared the estimation via Monte Carlo simulations in terms of their mean and standard deviations to gain an idea on the estimation methods.

4.1 Conditional least square estimation

The conditional least squares estimators of the parameters α and λ are obtained by minimizing the function

$$Q_n = \sum_{t=2}^n (X_t - \alpha X_{t-1} - (1 - \alpha)\mu)^2$$

where $\mu = \frac{1}{e^\lambda - 1}$. The estimators are given by

$$\hat{\alpha}_{CLS} = \frac{(n-1) \sum_{t=2}^n (X_{t-1} X_t) - \sum_{t=2}^n X_t \sum_{t=2}^n X_{t-1}}{(n-1) \sum_{t=2}^n X_{t-1}^2 - \left(\sum_{t=2}^n X_{t-1}\right)^2}, \quad \hat{\mu}_{CLS} = \frac{\sum_{t=2}^n X_t - \hat{\alpha}_{CLS} \sum_{t=2}^n X_{t-1}}{(n-1)(1 - \hat{\alpha}_{CLS})}$$

and the estimator for parameter λ is $\hat{\lambda}_{CLS} = \ln\left(\frac{1+\hat{\mu}_{CLS}}{\hat{\mu}_{CLS}}\right)$.

Now we derive the asymptotic properties of the estimators. Since the process $\{X_t\}$ is strictly stationary and ergodic, it follows that the estimators $\hat{\alpha}_{CLS}, \hat{\lambda}_{CLS}$ are strongly consistent estimators for the parameters α, λ and $\sqrt{n}\left(\hat{\alpha}_{CLS} - \alpha, \hat{\lambda}_{CLS} - \lambda\right)'$ converges in distribution to the bivariate normal distribution with mean zero vector and covariance matrix given by

$$\begin{bmatrix} \frac{(1-\alpha)[\alpha(e^{2\lambda}-1)+e^\lambda(\alpha+1)]}{e^\lambda} & \alpha \\ \alpha & \frac{(1+\alpha)e^\lambda}{(1-\alpha)(e^\lambda-1)^2} \end{bmatrix}$$

4.2 Non-Parametric estimation

Consider the sample auto covariance and mean of X_t

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t-k} - \bar{X}), 0 \leq k < n, \quad \bar{X} = \frac{1}{n} \sum_{t=1}^n X_t.$$

Since $\alpha = \frac{\gamma(1)}{\gamma(0)}$ and $\mu = E(X_t)$, the non-parametric estimators of α and μ are

$$\hat{\alpha}_{NP} = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{\sum_{t=2}^n (X_t - \bar{X})(X_{t-1} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}, \quad \hat{\mu}_{NP} = \bar{X} = \frac{1}{e^{\hat{\lambda}} - 1}$$

So the estimator for parameter λ is $\hat{\lambda}_{NP} = \ln\left(\frac{1+\bar{X}}{\bar{X}}\right)$.

4.3 Maximum likelihood estimation

The maximum likelihood estimation of the parameters are obtained by maximization of the log-likelihood function

$$\ln L(X_1, \dots, X_n; \alpha, \lambda) = -\lambda x_1 + \ln(1 - e^{-\lambda}) + \sum_{l=1}^n \ln p(X_l | X_{l-1}; \alpha, \lambda)$$

where $p(X_l | X_{l-1}; \alpha, \lambda)$ is given by (3.1). The MLEs can be easily computed by using the function `nlm` from statistical package R, taking the CLS estimators as initial values of the function `nlm`.

5 Simulation comparison

The performance of the CLS, NP and ML estimators is checked by Monte Carlo simulation using different sample sizes ($n = 100, 500, 1000, 5000$), where 1000 samples are simulated from the DEX-PINAR(1) process. Based on this simulation, Figure 1 displays some sample paths of DEXPINAR(1)

process with different choices of the parameter. Table 1 gives the mean and standard deviation (in brackets) of the estimates for different values of the parameters with different sample sizes. Based on this table, we find that the estimates obtained from the three estimation methods are convergent in their values. Also, increasing the sample size implies smaller standard deviation and the MLEs converge faster to the true values of the parameters. Moreover, we can conclude that the MLEs have the smallest standard deviations than the others and hence the MLEs provide the best performance.

Table 1: Some simulation results for the estimation of the parameters α, λ with standard errors in brackets						
n	$\hat{\alpha}_{ML}$	$\hat{\alpha}_{CLS}$	$\hat{\alpha}_{NP}$	$\hat{\lambda}_{ML}$	$\hat{\lambda}_{CLS}$	$\hat{\lambda}_{NP}$
100	0.3994309 (0.0188353)	0.396499 (0.0333889)	0.398963 (0.0366721)	0.5002217 (0.0218871)	0.5008763 (0.0261026)	0.5014786 (0.0229565)
500	0.4000712 (0.0165113)	0.3967013 (0.0327761)	0.3990146 (0.0336290)	0.5001608 (0.0192652)	0.5002066 (0.0248352)	0.5002805 (0.0217358)
1000	0.3999713 (0.0061732)	0.3996258 (0.0099159)	0.3998021 (0.0082805)	0.5000232 (0.0032049)	0.5000841 (0.0064249)	0.5000688 (0.0083846)
5000	0.399999 (0.0002357)	0.3999807 (0.0025514)	0.3999381 (0.0012267)	0.5000016 (0.0009274)	0.5000136 (0.0026035)	0.5000045 (0.0024501)
$(\alpha, \lambda) = (0.8, 3)$						
100	0.7965457 (0.0571306)	0.7890544 (0.0700193)	0.7925141 (0.0617926)	2.931672 (0.3891565)	3.111916 (0.5285243)	3.080586 (0.3575966)
500	0.799149 (0.0362241)	0.7971213 (0.0631412)	0.7988763 (0.0407399)	2.982694 (0.2195496)	3.0455832 (0.3024964)	2.99397 (0.2815863)
1000	0.8002674 (0.0037831)	0.7998051 (0.0071153)	0.7993048 (0.0061474)	2.992536 (0.1029734)	2.9823748 (0.1954307)	3.004411 (0.1740586)
5000	0.80005372 (0.0007192)	0.8002186 (0.0017384)	0.7999751 (0.000946)	2.999923 (0.0063882)	2.992036 (0.0290423)	3.0007531 (0.0091607)

Conclusion

Discrete data can be modeled by new stationary time series model as DEXPINAR(1), which is over dispersion. The parameters of the model can be estimated by three method as: conditional least square, non parametric and maximum likelihood. The ML estimator gives the best result among the others.

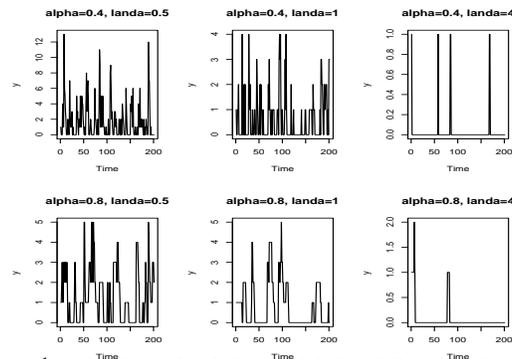


Figure 1: sample path of the model, for different value for α, λ .

References

- Al-Osh, M. A. and Alzaid, A. A. (1987) First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* **8**, 261-275.
- Aly, E. E.A.A., Bouzar, N. (1994) On Some Integer-Valued Autoregressive Moving Average Models, *Journal of Multivariate Analysis*, 50, 132-151.
- McKenzie E (1985) Some simple models for discrete variate time series. *Water Res Bull* **21**, 645-650.
- Ristić, M. M., Bakouch, H. S. and Nastić, A. S. (2009) A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *J. Stat. Plan. Inference* **139**, 2218-2226.
- Ristić, M. M., Nastić, A. S. and Bakouch, H. S. (2012) Estimation in an integer-valued autoregressive process with negative binomial marginals (NBINAR(1)), *Communications in Statistics: Theory and Methods* **41**, 606-618.
- Schweer, S. and Weiß, C.H. (2014). Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. *Computational Statistics and Data Analysis* **77**, 267-284.
- Stutel, FW, van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Ann Probab*, **7**, 893-899.



A new measure of inaccuracy based on past lifetime

R. Siroosi, and M. Khorashadizadeh¹⁰

Department of Statistics, School of Mathematical Sciences and Statistics,
University of Birjand, Iran.

Abstract: Kerridge (1961) has defined a measure of inaccuracy which is a generalization of Shannon's (1948) entropy. In literature of information science, several generalization and alternative measure of Kerridge inaccuracy and Shannon entropy have been studied. In this paper, we try to generalize Kerridge inaccuracy measure based on cumulative distribution of the past lifetime ($t - X|X \leq t$). We also obtain its relation with some reliability concepts such as reversed hazard rate (RHR), mean past lifetime (MPL) and proportional reversed hazard rate model (PRHM). A bound for the proposed measure in terms of MPL function has been derived. Finally, it is shown that the DCPI under some conditions, can determine the distribution function uniquely.

Keywords Shannon entropy, Kerridge inaccuracy, cumulative past inaccuracy, dynamic cumulative past inaccuracy, reversed hazard rate, mean past lifetime, proportional reversed hazard model.

Mathematics Subject Classification (2010): 62B10 62N05 90B25.

1 Introduction

Suppose X and Y are two non negative continuous random variables with probability density functions $f(t)$ and $g(t)$ respectively, and let $f(t)$ is the actual pdf corresponding to the observations and $g(t)$ is the pdf estimated by experimenter, then as a generalization of the Shannon entropy, Kerridge (1961) and Nath (1968) have introduced a useful tool for measurement of error in experimental results as follow,

$$\begin{aligned} H(X, Y) &= - \int_0^{\infty} f(x) \log g(x) dx \\ &= - \int_0^{\infty} f(x) \log f(x) dx + \int_0^{\infty} f(x) \log \frac{f(x)}{g(x)} dx, \end{aligned} \quad (1.1)$$

where the first item of (1.1) is the error of uncertainty, which is Shannon's entropy,

$$H(X) = - \int_0^{\infty} f(x) \log f(x) dx$$

¹⁰Speaker's Name : m.khorashadizadeh@birjand.ac.ir

and the second term is the Kullback-Leibler measure of discrimination (for more details about Kullback-Leibler measure see [Kullback \(1959\)](#)). If $f(x) = g(x)$ then $H(X, Y)$ achieves its minimum value, which is the Shannon's entropy, $H(X)$. For more properties and applications of the $H(X, Y)$ one can see the [Smitha \(2010\)](#) and the references there in.

In application, since the current age of the system under consideration is also taken into account, thus the two measures $H(X)$ and $H(X, Y)$ are not suitable. Therefore the concept of residual measures have been developed in the literature. Let X be a continuous random variable denoting the lifetime of a device or a system. Two dual random variables called residual lifetime, $X_t = (X - t | X \geq t)$, and past lifetime, $X_t^* = (t - X | X \leq t)$, are the fundamental concepts in reliability and information theory. [Ebrahimi and Pellerey \(1995\)](#) and [Ebrahimi \(1996\)](#) defined the residual entropy (RE) based on the random variable $X_t = (X - t | X > t)$ by,

$$\begin{aligned} RE(X; t) &= - \int_t^\infty \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)} dx, \\ &= 1 - E(\log \lambda(X) | X \geq t), \end{aligned}$$

where $\lambda(x) = \frac{f(x)}{F(x)}$ is the hazard rate function and $RE(X; 0) = H(X)$.

Accordingly, [Nair and Gupta \(2007\)](#) and [Taneja et al. \(2009\)](#) have extended the measure of inaccuracy, in dynamic version, given as

$$H(X, Y; t) = - \int_t^\infty \frac{f(x)}{\bar{F}(t)} \log \frac{g(x)}{\bar{G}(t)} dx, \quad (1.2)$$

where when $t \rightarrow 0$, it implies $H(X, Y)$. It can be easily seen that the relation among hazard rates of X and Y and $H(X, Y; t)$ is as follows,

$$\lambda_X(t) = \frac{\frac{\partial H(X, Y; t)}{\partial t} + \lambda_Y(t)}{H(X, Y; t) + \log \lambda_Y(t)}.$$

Recently, [Di Crescenzo and Longobardi \(2002\)](#) introduced an entropy-based measure of uncertainty in past life time distributions and called it past entropy. The past entropy (PE) represents the uncertainty of the idle time (or inactivity time) of a component or system which is based on past life time random variable, $X_t^* = (t - X | X \leq t)$, and it is given by,

$$\begin{aligned} PE(X; t) &= - \int_0^t \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} dx \\ &= 1 - E(\log \lambda^*(X) | X \leq t), \end{aligned}$$

where $\lambda^*(t) = \frac{f(x)}{F(x)}$ is the reversed hazard rate function and $PE(X; \infty) = H(X)$.

Nair and Gupta (2007) and Kumar et al. (2011), have studied a dynamic measure of inaccuracy based on the past entropy of the form,

$$H^*(X, Y; t) = - \int_0^t \frac{f(x)}{F(t)} \log \frac{g(x)}{G(t)} dx, \quad (1.3)$$

where $H^*(X, Y; \infty) = H(X, Y)$.

More recently, Rao et al. (2004) defined a new uncertainty measure, the cumulative residual entropy (CRE), through,

$$CRE(X) = - \int_0^\infty \bar{F}(x) \log \bar{F}(x) dx.$$

CRE has many good properties. First, its definition is valid in the continuous and discrete cases which is more general than the Shannon entropy; second, it has more general mathematical properties than the Shannon entropy; and at last, it can be easily estimated from sample data and the estimation asymptotically converge to the true values. One can see many properties and applications of CRE in Rao (2005), Liu (2007) and Di Crescenzo and Longobardi (2009).

Asadi and Zohrevand (2007) defined the dynamic version of cumulative residual entropy (DCRE), which is the CRE of the residual random variable X_t and it is given by,

$$\begin{aligned} DCRE(X; t) &= - \int_t^\infty \frac{\bar{F}(x)}{\bar{F}(t)} \log \frac{\bar{F}(x)}{\bar{F}(t)} dx, \\ &= - \frac{1}{\bar{F}(t)} \int_t^\infty \bar{F}(x) \log \bar{F}(x) dx + \mu(t) \log \bar{F}(t), \end{aligned}$$

where $\mu(t) = E(X - t | X \geq t)$ is the mean residual life (MRL) function and $DCRE(X; 0) = CRE(X)$.

As a generalization of the CRE and DCRE, Taneja and Kumar (2012) have defined the cumulative residual inaccuracy (CRI) and then the dynamic version of it by,

$$\begin{aligned} CRI(X, Y) &= - \int_0^\infty \bar{F}(x) \log \bar{G}(x) dx \\ &= - \int_0^\infty \bar{F}(x) \log \bar{F}(x) dx + \int_0^\infty \bar{F}(x) \log \frac{\bar{F}(x)}{\bar{G}(x)} dx, \end{aligned}$$

and

$$DCRI(X, Y; t) = - \int_t^\infty \frac{\bar{F}(x)}{\bar{F}(t)} \log \frac{\bar{G}(x)}{\bar{G}(t)} dx.$$

The cumulative past entropy (CPE) has also introduced and studied Di Crescenzo and Longobardi (2009) as follows,

$$CPE(X) = - \int_0^\infty F(x) \log F(x) dx. \quad (1.4)$$

Furthermore, Navarro et al. (2010) studied many properties of DCRE and defined the dynamic version of cumulative past entropy (DCPE) by,

$$\begin{aligned} DCPE(X; t) &= - \int_0^t \frac{F(x)}{F(t)} \log \frac{F(x)}{F(t)} dx \\ &= \mu^*(t) \log F(t) - \frac{1}{F(t)} \int_0^t F(x) \log F(x) dx, \end{aligned} \quad (1.5)$$

where $\mu^*(t) = E(t - X | X \leq t)$ is the mean past life (MPL) function and $DCPE(X; \infty) = CPE(X)$.

Another reliability measure that has important role in modeling and dependance structure among two distributions is the proportional (reversed) hazard model. Suppose that two random variables X and Y satisfy the proportional reversed hazard rate model (PRHM) with proportionality constant $\beta (> 0)$, if

$$\lambda_Y^*(t) = \beta \lambda_X^*(t); \quad \beta > 0.$$

So, the PRHM is equivalent to the model,

$$G(t) = [F(t)]^\beta, \quad (1.6)$$

where $F(x)$ is the baseline distribution function and $G(x)$ can be considered as some reference distribution function. This model was proposed by Gupta et al. (1998) in contrast to the celebrated proportional hazard model (PHM), $\lambda_Y(t) = \beta \lambda_X(t)$. Although, the PRHM and PHM are strongly similar, but only PRHM is useful in analyze the left censored or right truncated data.

In this paper, first we propose the cumulative past inaccuracy (CPI) and its dynamic version (DCPI), which is related to CRE and cumulative Kullback-Leibler information. Then we obtain an upper bound for DCPI based on mean past lifetime function. We also show that the DCPI measure uniquely can determine the distribution function. Finally its relation to mean past lifetime is implied.

2 Main results

We proposed the cumulative past inaccuracy (CPI) and its dynamic version as a generalization of CPE (1.4) and DCPE (1.5), respectively by,

$$\begin{aligned} CPI(X, Y) &= - \int_0^\infty F(x) \log G(x) dx \\ &= - \int_0^\infty F(x) \log F(x) dx + \int_0^\infty F(x) \log \frac{F(x)}{G(x)} dx, \end{aligned} \quad (2.1)$$

and

$$DCPI(X, Y; t) = - \int_0^t \frac{F(x)}{F(t)} \log \frac{G(x)}{G(t)} dx$$

$$= \mu^*(t) \log G(t) - \frac{1}{F(t)} \int_0^t F(x) \log G(x) dx. \tag{2.2}$$

It should be noted that,

$$CPI(X, Y) = CRE(X) + CKL(X, Y) + [E(Y) - E(X)],$$

where $CKL(X, Y) = \int_0^\infty F(t) \log \frac{F(t)}{G(t)} dt + [E(X) - E(Y)]$ is the cumulative Kullback-Leibler information introduced by Park et al. (2012) and Di Crescenzo and Longobardi (2015). Also we have,

$$DCPI(X, Y; t) = DCRE(X, Y; t) + DCKL(X, Y; t) + [\mu_X^*(t) - \mu_Y^*(t)],$$

where $DCKL(X, Y; t) = \int_0^t \frac{F(x)}{F(t)} \log \left(\frac{F(x)G(t)}{F(t)G(x)} \right) dx + [\mu_Y^*(t) - \mu_X^*(t)]$ is the dynamic cumulative Kullback-Leibler information (Di Crescenzo and Longobardi (2015)).

In the following, we obtain an upper bound for dynamic cumulative past inaccuracy which is useful tool in the next results.

Theorem 2.1. *Let X and Y be two non-negative random variables satisfying the proportional reversed hazard model (PRHM) of the form (1.6). Then $DCPI(X, Y; t)$ is increasing function of t , if and only if the following inequality holds for all t ;*

$$DCPI(X, Y; t) \leq \beta \mu^*(t). \tag{2.3}$$

Proof. Using (2.2) and properties of PRHM and differentiating both sides with respect to t , the required result is follow. □

In the next theorem, we show that under proportional reversed hazard model (1.6), the DCPI can characterizes the distribution function F .

Theorem 2.2. *If X and Y be two non-negative random variables satisfying the PRHM and $DCPI(X, Y; t)$ is increasing function of t , then $DCPI$ uniquely determines the distribution function F .*

Proof. Let X_1, Y_1 and X_2, Y_2 be two sets of random variables with distribution functions F_1, G_1 and F_2, G_2 respectively, which both of them satisfying the proportional reversed hazard rate model, that is,

$$\lambda_{Y_1}^*(t) = \beta \lambda_{X_1}^*(t)$$

and

$$\lambda_{Y_2}^*(t) = \beta \lambda_{X_2}^*(t).$$

For proving the theorem, we will show that if for all $t \geq 0$,

$$DCPI(X_1, Y_1; t) = DCPI(X_2, Y_2; t), \tag{2.4}$$

then we should have $\lambda_{X_1}^*(t) = \lambda_{X_2}^*(t)$, which leads to $F_{X_1}(t) = F_{X_2}(t)$. So, let

$$A = \{t : t \geq 0 \text{ and } \lambda_{X_1}^*(t) \neq \lambda_{X_2}^*(t)\},$$

and assume that the set A is not empty and without loss of generality suppose that for some $t_0 \in A$,

$$\lambda_{X_1}^*(t_0) > \lambda_{X_2}^*(t_0). \quad (2.5)$$

On the other hand, differentiating the both sides of (2.4) with respect to t , simplifying and using the relations (2.5) and (2.3), we have

$$\beta\mu_{X_1}^*(t_0) - DCPI(X_1, Y_1; t_0) < \beta\mu_{X_2}^*(t_0) - DCPI(X_2, Y_2; t_0).$$

Since, the last inequality implies a contradiction $\mu_{X_1}^*(t_0) < \mu_{X_2}^*(t_0)$, the set A should be empty and the proof completed. \square

In the following theorem we show that, the DCPI has a direct relation with mean past lifetime.

Theorem 2.3. *Let X and Y are two nonnegative continuous random variables satisfying the PRHM (1.6), such that $DCPI(X, Y; t) < \infty$ for all $t > 0$. Then,*

$$DCPI(X, Y; t) = \beta E(\mu^*(X)|X \leq t). \quad (2.6)$$

Proof. Under PRHM (1.6) and by showing that $E(\mu^*(X)|X \leq t) = DCPE(X; t)$, the proof is straightforward. \square

Conclusion

We considered some extensions of Kerridge inaccuracy measure due to the past lifetime and cumulative distribution function. The CPI and its dynamic version is defined and an upper bound is obtained. We also show that it can characterized the distribution uniquely and its relationships with other reliability concepts are obtained. Of course, other properties of these functions are still waiting to be discovered.

References

- Asadi, M. and Zohrevand, Y. (2007). On the dynamic cumulative residual entropy. *Journal of Statistical Planning and Inference*, 137, 6, 1931-1941.
- Di Crescenzo, A. and Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. *Journal Appl. Prob.*, 39, 434-440.

- Di Crescenzo, A. and Longobardi, M. (2009). On cumulative entropies and lifetime estimations. In *Methods and Models in Artificial and Natural Computation. A Homage to Professor Miras Scientific Legacy* (pp. 132-141). Springer Berlin Heidelberg.
- Di Crescenzo, A. and Longobardi, M. (2015). Some properties and applications of cumulative Kullback-Leibler information. *Applied Stochastic Models in Business and Industry*.
- Ebrahimi, N. (1996). How to measure uncertainty in the residual lifetime distribution. *Sankhya A*, 58, 48-56.
- Ebrahimi, N. and Pellerey, F. (1995). New partial ordering of survival functions based on notation of uncertainty. *J. Appl. Probab.*, 32, 202-211.
- Gupta, R.C., Gupta, P.L. and Gupta, R.D. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics-Theory and Methods*, 27, 4, 887-904.
- Kerridge, D.F. (1961). Inaccuracy and inference. *J. R. Stat. Soc. Ser. B*, 23, 184-194.
- Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York.
- Kumar, V., Taneja, H.C. and Srivastava, R. (2011). A dynamic measure of inaccuracy between two past lifetime distributions. *Metrika*, 74, 1, 1-10.
- Liu, J. (2007). *Information theoretic content and probability*. Doctoral dissertation, University of Florida.
- Nair, N.U. and Gupta, R.P. (2007). Characterization of proportional hazard models by properties of information measures. *International Journal of Statistical Sciences*, 6, 223-231.
- Nath, P. (1968). Inaccuracy and coding theory. *Metrika*, 13, 1, 123-135.
- Navarro, J., del Aguila, Y. and Asadi, M. (2010). Some new results on the cumulative residual entropy. *Journal of Statistical Planning and Inference*, 140, 1, 310-322.
- Park, S., Rao, M. and Shin, D.W. (2012). On cumulative residual Kullback-Leibler information. *Statistics & Probability Letters*, 82, 11, 2025-2032.
- Rao, M. (2005). More on a new concept of entropy and information. *J. Theoretical Probability*, 18, 967-981.
- Rao, M., Chen, Y., Vemuri, B.C. and Wang, F. (2004). Cumulative residual entropy: a new measure of information. *IEEE Transactions on Information Theory*, 50, 6, 1220-1228.

- Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27, 379-423.
- Smitha, S. (2010). A Study on the Kerridges Inaccuracy Measure and Related Concepts. Doctoral dissertation, Cochin University of Science and Technology.
- Taneja, H.C. and Kumar, V. (2012). On Dynamic Cumulative Residual Inaccuracy Measure. In *Proceedings of the World Congress on Engineering (London, U.K.)*, Vol. 1.
- Taneja, H.C., Kumar, V. and Srivastava, R. (2009). A dynamic measure of inaccuracy between two residual lifetime distributions. *Int. Math. Fourm*, 4, 25, 1213-1220.



On bivariate dynamic cumulative past entropy

R. Siroosi¹, and M. Khorashadizadeh

Department of Statistics, School of Mathematical Sciences and Statistics,
University of Birjand, Iran.

Abstract: In this paper, we define two version of dynamic cumulative past entropy (DCPE) in bivariate setup and study their relations with reliability concepts such as bivariate reversed hazard rate (BRHR) and bivariate reversed mean residual lifetime (BRMRL). It is shown that the bivariate DCPE is not invariant under non-singular transformations. We also obtain an upper bound for bivariate DCPE based on BRMRL. Finally an example due to standard log-logistic distribution is presented.

Keywords Shannon entropy, cumulative past entropy, dynamic cumulative past entropy, bivariate reversed hazard rate, bivariate reversed mean residual lifetime.

Mathematics Subject Classification (2010): 62B10 62N05 90B25.

1 Introduction

The main measure of the uncertainty contained in non-negative random variable X with an absolutely continuous distribution function $F(x)$ and probability density function (PDF) $f(x)$ is the [Shannon's \(1948\)](#) entropy defined by,

$$H(X) = - \int_0^{\infty} f(x) \log f(x) dx.$$

After Shannon entropy several measure of uncertainty in different situations have been defined. [Di Crescenzo and Longobardi \(2002\)](#) defined dynamic past entropy for the random variable $X_t = (t - X | X \leq t)$, which $t \geq 0$, by

$$\bar{H}(t) = - \int_0^t \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} dx,$$

where $\frac{f(x)}{F(t)}$ is the PDF of X_t and $\bar{H}(t) \in [-\infty, \infty]$. Given that a device has been failed before time t , $\bar{H}(t)$ measures the uncertainty about it's past lifetime. Although Shannon's entropy has applications in many bases of researches, but [Rao et al. \(2004\)](#) introduced cumulative residual entropy (CRE) for the random lifetime X as follow,

$$\xi(X) = - \int_0^{\infty} \bar{F}(t) \log \bar{F}(t) dx,$$

¹Speaker's Name : siroosi.razieh89@yahoo.com

which has more mathematical properties and is related to reliability measures and also its estimation is more easier and accurate than Shannon entropy. Based on CRE, Di Crescenzo and Longobardi (2009), defined cumulative past entropy (CPE) for the past lifetime X , through the relationship,

$$\xi_c(X) = - \int_0^\infty F(t) \log F(t) dx. \quad (1.1)$$

Di Crescenzo and Longobardi (2009) and Navarro et al. (2010) have extended the concept of CPE in terms of conditional measure called the dynamic cumulative past entropy (DCPE). For a non-negative random variable X , which is the past lifetime of a component, the DCPE is the CPE associated with the random variable X_t , with the distribution function $F_t(x) = \frac{F(x)}{F(t)}$, and is defined as

$$\xi_c(X; t) = - \int_0^t F_t(x) \log F_t(x) dx, \quad (1.2)$$

where $t > 0$, $F(t) > 0$. We know that in univariate case, the reversed mean residual lifetime is as follow,

$$m(x) = \frac{1}{F(x)} \int_0^x F(u) du.$$

In real environments, systems are composed of two or more components that depend on each other for life. Therefore, we must examine the concepts of reliability in these circumstances and achieve their features and capabilities. Despite of univariate case, only few works have been done in higher dimensions. The works of Ahmad and Gokhale (1989), Zografos (1999), Darbellay and Vajda (2000), Rajesh and Nair (2000), Nadarajah and Zografos (2005), Ebrahimi et al. (2007), Rajesh et al. (2009), Sathar et al. (2009, 2010) and Khorashadizadeh et al. (2013) focus attention on extending information measures in multi-component systems. Accordingly, Roy (2002) defined bivariate reversed hazard rate (BRHR) of a random vector $\underline{X} = (X_1, X_2)$ as a two component vector given by,

$$\underline{rh}(x_1, x_2) = (rh_1(x_1, x_2), rh_2(x_1, x_2)), \quad (1.3)$$

where

$$\begin{aligned} rh_i(x_1, x_2) &= \lim_{\Delta x_i \rightarrow 0} \frac{P(x_i - \Delta x_i | X_1 \leq x_1, X_2 \leq x_2)}{\Delta x_i} \\ &= \frac{\partial \log F(x_1, x_2)}{\partial x_i}, \quad i = 1, 2. \end{aligned} \quad (1.4)$$

For $i = 1$, $rh_1(x_1, x_2)\Delta x_1$ is the probability of failure of the first component in the interval $(x_1 - \Delta x_1, x_1]$ given that it has failed before x_1 and the second has failed before x_2 . The interpretation for $rh_2(x_1, x_2)$ is similar. Also, Bisimi (2005) has been defined and studied a scalar BRHR as

$$r(x_1, x_2) = \frac{f(x_1, x_2)}{F(x_1, x_2)}.$$

Similarly to vector BRHR stated in (1.3), Nair and Asha (2008) defined bivariate reversed mean residual life (BRMRL) function as the following vector,

$$\underline{m}(x_1, x_2) = (m_1(x_1, x_2), m_2(x_1, x_2)), \tag{1.5}$$

where

$$\begin{aligned} m_i(x_1, x_2) &= E(x_i - X_i | X_1 \leq x_1, X_2 \leq x_2) \\ &= \frac{1}{F(x_1, x_2)} \int_0^{x_i} F(x_i, x_j) dx_i, \quad i, j = 1, 2; i \neq j. \end{aligned} \tag{1.6}$$

They also have been shown that the BRMRL and the BRHR have a relation as follow,

$$r h_i(x_1, x_2) = (m_i(x_1, x_2))^{-1} \left(1 - \frac{\partial m_i(x_1, x_2)}{\partial x_i} \right), \quad i = 1, 2.$$

In the rest of the paper, we define a vector and scalar versions of bivariate dynamic CPE and obtain some of their properties. We show that our proposed measure is invariant under non-singular transformations and its relation with reversed mean residual function is obtained. Also, an upper bound based on BRMRL is presented.

2 Definitions and properties

In this section we try to extend CPE and DCPE to the bivariate setup. A natural extension of (1.1) and (1.2) is given in next definitions.

Definition 2.1. Let $\underline{X} = (X_1, X_2)$ be a non-negative bivariate random vector with an absolutely continuous distribution function $F(x_1, x_2)$, probability density function $f(x_1, x_2)$ and marginal distribution functions $F_i(x_i), i = 1, 2$. Then the scalar bivariate cumulative past entropy (BCPE) can be defined through the relationship

$$BC(X_1, X_2) = - \int_0^\infty \int_0^\infty F(x_1, x_2) \log F(x_1, x_2) dx_1 dx_2. \tag{2.1}$$

Remark 2.2. If X_1 and X_2 are independent, then (2.1) is equal to

$$\begin{aligned} BC(X_1, X_2) &= - \int_0^\infty F_1(x_1) \log F_1(x_1) dx_1 - \int_0^\infty F_2(x_2) \log F_2(x_2) dx_2, \\ &= \xi_c(X_1) + \xi_c(X_2) \end{aligned}$$

and also if X_1 and X_2 have same distribution, then we have

$$BC(X_1, X_2) = 2\xi_c(X_1).$$

Definition 2.3. Let $\underline{X} = (X_1, X_2)$ be a non-negative bivariate random vector which have the same properties of Definition 2.1, Then the scalar bivariate dynamic cumulative past entropy (BDCPE) can be defined as follow

$$BDC(t_1, t_2) = - \int_0^{t_1} \int_0^{t_2} \frac{F(x_1, x_2)}{F(t_1, t_2)} \log \frac{F(x_1, x_2)}{F(t_1, t_2)} dx_1 dx_2. \quad (2.2)$$

If $\underline{X} = (X_1, X_2)$ be the failure time of a two-component system, (2.2) can be considered as a measure of uncertainty of failure time when the components have been found down at time $\underline{t} = (t_1, t_2)$.

Remark 2.4. Note that $BDC(\infty, \infty) = BC(X_1, X_2)$. So the properties of BDCPE that will be said in the following are true for BCPE.

Rajesh et al. (2014) have shown that bivariate dynamic cumulative residual entropy (BDCRE) is not influenced by non-singular transformations. Similarly we show that BDCPE defined in (2.2) is not invariant under non-singular transformations.

Remark 2.5. If $Y_j = \Phi_j(X_j)$, $j = 1, 2$, are one-to-one transformations with joint distribution function $G(y_1, y_2)$ and bivariate dynamic CPE, $BDC(t_1, t_2)$, then

$$BDC(\Phi_1(t_1), \Phi_2(t_2)) = - \int_0^{t_1} \int_0^{t_2} \frac{F(x_1, x_2)}{F(t_1, t_2)} \log \frac{F(x_1, x_2)}{F(t_1, t_2)} J dx_1 dx_2,$$

where $J = \left| \frac{\partial}{\partial x_1} \Phi_1(x_1) \frac{\partial}{\partial x_2} \Phi_2(x_2) \right|$ is the absolute value of the Jacobian of transformation. The above expression represents that (2.2) is not invariant under non-singular transformations. In the specific case, if we choose $\Phi_j(X_j) = a_j x_j + b_j$, then we have

$$BDC(\Phi_1(t_1), \Phi_2(t_2)) = a_1 a_2 BDC(t_1, t_2).$$

Remark 2.6. If X_1 and X_2 are independent, then

$$\begin{aligned} BDC(t_1, t_2) &= - \int_0^{t_1} \int_0^{t_2} \frac{F_1(x_1)}{F_1(t_1)} \frac{F_2(x_2)}{F_2(t_2)} \log \frac{F_1(x_1)}{F_1(t_1)} \frac{F_2(x_2)}{F_2(t_2)} dx_1 dx_2 \\ &= - \int_0^{t_2} \frac{F_2(x_2)}{F_2(t_2)} dx_2 \int_0^{t_1} \frac{F_1(x_1)}{F_1(t_1)} \log \frac{F_1(x_1)}{F_1(t_1)} dx_1 \\ &\quad - \int_0^{t_1} \frac{F_1(x_1)}{F_1(t_1)} dx_1 \int_0^{t_2} \frac{F_2(x_2)}{F_2(t_2)} \log \frac{F_2(x_2)}{F_2(t_2)} dx_2 \\ &= m_2(t_2) \xi_c(X_1; t_1) + m_1(t_1) \xi_c(X_2; t_2), \end{aligned}$$

where $m_i(x_i)$ and $\xi_c(X_i; t_i)$ are the reversed mean residual life function and the DCPE of X_i , $i = 1, 2$, respectively. In particular, if X_1 and X_2 are independent and have the same reversed mean residual life function $m(t)$, the above relation corresponds

$$BDC(t_1, t_2) = m(t) (\xi_c(X_1; t_1) + \xi_c(X_2; t_2)).$$

Now, we investigate the behavior of the DCPE for conditional distributions. Consider the random variables $Y_j = [X_j|X_i < t_i], i = 1, 2$. The distribution of Y_j corresponds the conditional distribution of X_j provided that X_i has been failed to time $t_i, i = 1, 2$, and has distribution function $\frac{F(x_1, t_2)}{F(t_1, t_2)}$ for $x_1 < t_1$ and $\frac{F(t_1, x_2)}{F(t_1, t_2)}$ for $x_2 < t_2$. The DCPE for the random variables $Y_j, j = 1, 2$ is equal to

$$BDC_1(t_1, t_2) = - \int_0^{t_1} \frac{F(x_1, t_2)}{F(t_1, t_2)} \log \frac{F(x_1, t_2)}{F(t_1, t_2)} dx_1, \tag{2.3}$$

and

$$BDC_2(t_1, t_2) = - \int_0^{t_2} \frac{F(t_1, x_2)}{F(t_1, t_2)} \log \frac{F(t_1, x_2)}{F(t_1, t_2)} dx_2. \tag{2.4}$$

Analogous to bivariate reversed hazard rate as a two component vector, we give another definition of vector BDCPE.

Definition 2.7. For a non-negative bivariate random vector $\underline{X} = (X_1, X_2)$ with an absolutely continuous distribution function $F(x_1, x_2)$, vector BDCPE is defined as

$$\underline{BDC}(t_1, t_2) = (BDC_1(t_1, t_2), BDC_2(t_1, t_2)). \tag{2.5}$$

where $BDC_1(t_1, t_2)$ and $BDC_2(t_1, t_2)$ are given by (2.3) and (2.4), respectively.

If $\underline{X} = (X_1, X_2)$ is a bivariate random vector, then $BDC_1(t_1, t_2)$ measures expected uncertainty contained in random variable X_1 about the predictability of the past lifetime of the first component, before t_1 , subject to the revision that X_2 has been failed before time t_2 . This interpretation can be said for $BDC_2(t_1, t_2)$, similarly. According to BRML function of the form (1.5), (2.3) and (2.4) can be rewritten as

$$BDC_1(t_1, t_2) = m_1(t_1, t_2) \log F(t_1, t_2) - \int_0^{t_1} \frac{F(x_1, t_2)}{F(t_1, t_2)} \log F(x_1, t_2) dx_1, \tag{2.6}$$

and

$$BDC_2(t_1, t_2) = m_2(t_1, t_2) \log F(t_1, t_2) - \int_0^{t_2} \frac{F(t_1, x_2)}{F(t_1, t_2)} \log F(t_1, x_2) dx_2, \tag{2.7}$$

where $m_i(t_1, t_2), i = 1, 2$ is said in (1.6).

Now, we compute the vector BDCPE which is defined above for the standard bivariate log-logistic distribution.

Example 2.8. Let (X, Y) have standard bivariate log-logistic distribution with CDF as follow,

$$F(x, y) = \left(1 + \frac{1}{xy} + \frac{1}{x} + \frac{1}{y} \right)^{-1}, \quad x, y > 0.$$

By using (2.3) and (2.4) and simplifying, for $i = 1, 2$ we get

$$BDC_i(t_1, t_2) = -\frac{1}{6} \left(\frac{t_i + 1}{t_i} \right) \left(\pi^2 + 6(\ln(t_i) - 3)\ln(t_i + 1) - 6 \operatorname{dilog} \left(\frac{t_i}{t_i + 1} \right) \right),$$

where $\operatorname{dilog}(t) = \int_1^t \frac{\ln(u)}{1-u} du$ is the dilogarithm function.

In the following theorem, we obtain an upper bound for $BDC_i(t_1, t_2)$, $i = 1, 2$, which has been defined in (2.6) and (2.7).

Theorem 2.9. *If $BDC_i(t_1, t_2)$ be increasing with respect to t_i , $i = 1, 2$. Then*

$$BDC_i(t_1, t_2) < m_i(t_1, t_2), \quad i = 1, 2. \quad (2.8)$$

Proof. For $i = 1$, by differentiating from (2.6) with respect to t_1 , we have

$$\begin{aligned} \frac{\partial}{\partial t_1} BDC_1(t_1, t_2) &= \frac{\partial m_1(t_1, t_2)}{\partial t_1} \log F(t_1, t_2) + m_1(t_1, t_2) \frac{\partial \log F(t_1, t_2)}{\partial t_1} \\ &\quad - \log F(t_1, t_2) + \frac{\frac{\partial F(t_1, t_2)}{\partial t_1}}{F^2(t_1, t_2)} \int_0^{t_1} F(x_1, t_2) \log F(x_1, t_2) dx_1. \end{aligned}$$

Now, by differentiating from (1.6) with respect to t_1 , using (1.4) and (2.6) and simplifying, we get

$$\frac{\partial}{\partial t_1} BDC_1(t_1, t_2) = rh_1(t_1, t_2)(m_1(t_1, t_2) - BDC_1(t_1, t_2)).$$

Knowing that $rh_1(t_1, t_2) > 0$, the proof is completed. The proof is similar for $i = 2$. \square

In the following theorem we obtain a functional relationship between $BDC(t_1, t_2)$ and the values of vector $\underline{m}(t_1, t_2)$.

Theorem 2.10. *Let $\underline{X} = (X_1, X_2)$ be a non-negative bivariate random vector with an absolutely continuous distribution function $F(t_1, t_2)$ with $BDC_i(t_1, t_2) < \infty$ and BRMRL function with components $m_i(t_1, t_2)$, $i = 1, 2$ for all $t_i \geq 0$. Then*

$$BDC_1(t_1, t_2) = \int_0^{t_1} \frac{f(x_1, t_2)}{F(t_1, t_2)} m_1(x_1, t_2) dx_1,$$

$$BDC_2(t_1, t_2) = \int_0^{t_2} \frac{f(t_1, x_2)}{F(t_1, t_2)} m_2(t_1, x_2) dx_2.$$

Proof. The proof concludes easily from (2.6) and (2.7). \square

References

- Ahmad, N.A. and Gokhale, D.V. (1989). Entropy expressions and their estimators for multivariate distributions. *IEEE Trans. Inform. Theory*, 35, 688-692.
- Bisimi, G. (2005). Bivariate Burr distributions. Unpublished PHD Thesis. Cochin University of Science and Technology.
- Darbellay, G.A. and Vajda, I. (2000). Entropy expressions for multivariate continuous distributions. *IEEE Trans. Inform. Theory*, 46, 709-712.
- Di Crescenzo, A. and Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. *Journal Appl. Prob.*, 39(2), 434-440.
- Di Crescenzo, A. and Longobardi, M. (2009). On cumulative entropies. *Journal of Statistical Planning and Inference*, 139, 4072-4087.
- Ebrahimi, N., Kirmani, S.N.U.A. and Soofi, E.S. (2007). Multivariate dynamic information. *J. Multivariate Anal.*, 98, 328-349.
- Khorashadizadeh, M., Rezaei Roknabadi, A.H. and Mohtashami Borzadaran, G.R. (2013). Doubly truncated (interval) cumulative residual and past entropy. *Statistics and Probability Letters*, 83(5), 1464-1471.
- Nadarajah, S. and Zografos, K. (2005). Expressions for Renyi and Shannon entropies for bivariate distributions. *Inform. Sci.*, 170, 173-189.
- Nair, N.U., Asha, G. (2008). Some characterizations based on bivariate reversed mean residual life. *Probst Forum*, 1, 1-14.
- Navarro, J., Aguila, Y., Asadi, M. (2010). Some new results on the cumulative residual entropy. *Journal of Statistical Planning and Inference*, 140, 310-322.
- Rajesh, G., Nair, K.R.M. (2000). Residual entropy of conditional distributions. *Stat. Methods*, 2, 72-80.
- Rajesh, G., Sathar, A.E.I., Nair, K.R.M. (2009). Bivariate extension of residual entropy and some characterization results. *J. Indian Statist. Assoc.*, 47, 91-107.
- Rajesh, G., Sathar, A.E.I., Nair, K.R.M., Reshmi, K.V. (2014). Bivariate extension of dynamic cumulative residual entropy. *Statistical Methodology*, 16, 72-82.

- Rao, M., Chen, Y., Vemuri, B.C. and Wang, F. (2004). Cumulative residual entropy: a new measure of information. *IEEE Transactions on Information Theory*, 50(6), 1220-1228.
- Roy, D. (2002). A characterization of model approach for generating bivariate life distributions using reversed hazard rates. *Journal of Japan Statistical Society*, 32(2), 239-245.
- Sathar, A.E.I., Nair, K.R.M. and Rajesh, G. (2009). Generalized bivariate residual entropy function and some characterization results. *South African Statist. J.*, 44, 1-18.
- Sathar, A.E.I., Rajesh, G. and Nair, K.R.M. (2010). Bivariate geometric vitality function and some characterization results. *Calcutta Statist. Assoc. Bull.*, 62, 247-267.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27, 279-423.
- Zografos, K. (1999). On maximum entropy characterization of Pearsons type II and VII multivariate distributions. *J. Multivariate Anal.*, 71, 67-75.



Generalized Mixture Representations for Inactivity Times of Coherent Systems

A. Sotoudeh¹, S. Goliforushani²

^{1,2} Department of Mathematical Sciences, Isfahan University of Technology, Isfahan

Abstract: In this paper, we derive some generalized mixture representations for the reliability functions of inactivity times of coherent systems under two specific conditions concerning the status of the components or the system in terms of reliability functions of inactivity times of order statistics.

Keywords Coherent system, Inactivity time, Order statistics, Signature; Reliability function, Generalized mixture.

Mathematics Subject Classification (2010): 62F15 62C10 62N05 65C60 62F10.

1 Introduction

The usefulness of signatures is evident from the mixture representation for the system lifetime in [samaniago \(1985\)](#). It shows that the lifetime distribution of a coherent system based on n components with i.i.d. lifetimes and common continuous distribution F can be expressed as a function which depends on the system's design only through its signature. Consider a coherent system comprising n components with independent and identically distributed (i.i.d.) lifetimes X_1, X_2, \dots, X_n distributed according to the common continuous distribution F . Let T be the system's lifetime. Then, the signature of the coherent system is defined to be a probability vector $\mathbf{s} = (s_1, s_2, \dots, s_n)$ such that

$$s_i = P(T = X_{i:n}), \quad i = 1, 2, \dots, n, \quad (1.1)$$

where $X_{i:n}$ denotes the i th ordered lifetime of the components. It is known that the signature vector \mathbf{s} does not depend on the underlying distribution function F . The following theorem, due to [samaniago \(1985\)](#), shows that the reliability function of the coherent system can be written as a mixture of the reliability functions of the ordered lifetimes of the components as follows.

Theorem 1.1. *If $\bar{F}_T(t)$ denotes the reliability function of the coherent system. Then,*

$$\bar{F}_T(t) = \sum_{i=1}^n s_i \bar{F}_{i:n}(t), \quad (1.2)$$

¹A. Sotoudeh : abdorrahim.sotoudeh@math.iut.ac.ir

where $\bar{F}_{i:n}(t)$ denotes the reliability function of $X_{i:n}$.

Signature vectors have been found to be quite useful in evaluating and comparing the performance of competing systems; see, for example, Kochar et al. (1999), Navarro et al. (2008), Navarro et al. (2005), Khaledi et al. (2007), Li et al. (2008), Zhang (2010), Goliforushani et al. (2012), and Goliforushani et al. (2011). Another mixture representation of the reliability function $\bar{F}_T(t)$ of the coherent system is as given in the following theorem.

Theorem 1.2. *Let T be the lifetime of a coherent system of order n with i.i.d. lifetimes X_1, X_2, \dots, X_n distributed according to a common underlying continuous distribution F . Then, the following generalized mixture representation of the reliability function of the system holds:*

$$\bar{F}_T(t) = \sum_{i=1}^n a_i \bar{F}_{1:i}(t), \quad (1.3)$$

where $\bar{F}_{1:i}(t) = P(X_{1:i} > t)$ is the reliability function of the series system with i components and the vector of coefficients $\mathbf{a} = (a_1, \dots, a_n)$ satisfying $\sum_{i=1}^n a_i = 1$ is referred to as the domination vector.

In Section 2, we focus on coherent systems under some specific conditions concerning the status of the components or the system and derive a generalized mixture representation for the inactivity time of the system.

2 Generalized mixture representation for inactivity times of coherent systems under some conditions

Consider a coherent system with lifetime T that has failed by time t , i.e., $\{T < t\}$, and suppose we know that at least $(n - i + 1)$ of the components are still alive, i.e., $\{X_{i:n} > t\}$. In this case, we can define the following conditional random variable :

$$(t - T | T < t, X_{i:n} > t). \quad (2.1)$$

To obtain a generalized mixture representation for 2.1, we first need the following lemma.

Lemma 2.1. *Let T be the lifetime of a coherent system and i ($1 < i \leq n$) be an integer. Then,*

$$P(T < y, X_{i:n} > x) = \sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k} \bar{F}^j(x) F^k(y) \quad \text{for all } y \leq x, \quad (2.2)$$

where the coefficients $a_{j,k}$ are real numbers that do not depend on F .

Proof. It is well known that the lifetime T of a coherent system can be written as $T = \min_{j=1, \dots, s} \max_{i \in Q_j} X_i$, where Q_1, Q_2, \dots, Q_s are minimal cut sets of the system. We use the following notation : $X_P = \min_{i \in P} X_i$ and $X^Q = \max_{i \in Q} X_i$. From the inclusion-exclusion formula, we have

$$\begin{aligned}
 P(T < y, X_{i:n} > x) &= P\left(\min_{j=1, \dots, s} \max_{i \in Q_j} X_i < y, X_{i:n} > x\right) \\
 &= P\left(\left(\bigcup_{j=1}^s \{X^{Q_j} < y\}\right) \cap \{X_{i:n} > x\}\right) \\
 &= P\left(\bigcup_{j=1}^s \left(\{X^{Q_j} < y\} \cap \{X_{i:n} > x\}\right)\right) \\
 &= \sum_{j=1}^s P(X^{Q_j} < y, X_{i:n} > x) - \sum_{j < k} P(X^{Q_j \cup Q_k} < y, X_{i:n} > x) \\
 &\quad + \dots \pm P(X^{Q_1 \cup Q_2 \cup \dots \cup Q_s} < y, X_{i:n} > x).
 \end{aligned}
 \tag{2.3}$$

Analogously, $X_{i:n} = \max_{j=1, \dots, m} \min_{i \in P_j} X_i$, where $m = \binom{n}{n-i+1}$ and P_1, \dots, P_m are all the subsets of $\{1, 2, \dots, n\}$ with exactly $(n-i+1)$ elements. Therefore, if $I \subset \{1, 2, \dots, n\}$ such that $|I| \leq i-1$ (note that $P(X^I < y, X_{i:n} > x) = 0$ if $|I| > i-1$), where $|I|$ denotes the cardinality of I , then

$$\begin{aligned}
 P(X^I < y, X_{i:n} > x) &= P\left(X^I < y, \max_{j=1, \dots, m} \min_{i \in P_j} X_i > x\right) \\
 &= P\left(\left(\bigcup_{j=1}^m \{X_{P_j} > x\}\right) \cap \{X^I < y\}\right) \\
 &= P\left(\bigcup_{j=1}^m \left(\{X_{P_j} > x\} \cap \{X^I < y\}\right)\right) \\
 &= \sum_{j=1}^m P(X_{P_j} > x, X^I < y) - \sum_{j < k} P(X_{P_j \cup P_k} > x, X^I < y) \\
 &\quad + \dots \pm P(X_{P_1 \cup P_2 \cup \dots \cup P_m} > x, X^I < y). \\
 &= \sum_{j=1}^{m^*} P(X_{P_j^*} > x, X^I < y) - \sum_{j < k} P(X_{P_j^* \cup P_k^*} > x, X^I < y) \\
 &\quad + \dots \pm P(X_{P_1^* \cup P_2^* \cup \dots \cup P_{m^*}^*} > x, X^I < y),
 \end{aligned}
 \tag{2.4}$$

where $m^* = \binom{|\bar{I}|}{n-i+1}$ and $P_1^*, \dots, P_{m^*}^*$ are all the subsets of $\bar{I} = \{1, 2, \dots, n\} - I$ with exactly $(n-i+1)$ elements. Now, upon using the fact that the components are i.i.d., from (8) we obtain

$$P(X^I < y, X_{i:n} > x) = c_{n-i+1} \bar{F}^{n-i+1}(x) F^{|\bar{I}|}(y) + c_{n-i+2} \bar{F}^{n-i+2}(x) F^{|\bar{I}|}(y) + \dots + c_{|\bar{I}|} \bar{F}^{|\bar{I}|}(x) F^{|\bar{I}|}(y),$$

where $c_{n-i+1}, c_{n-i+2}, \dots, c_{|\bar{I}|}$ are coefficients that do not depend on F . Using this last expression and 2.3, we obtain 2.2 upon grouping together terms with the same powers of $\bar{F}(x)$ and $F(y)$. □

We now present the main result of this section for the conditional random variable in 2.1.

Theorem 2.2. *Let T be the lifetime of a coherent system with i.i.d. components having a common continuous distribution function F and $i \in \{2, \dots, n\}$ such that $P(T < t, X_{i:n} > t) > 0$. Then, there exist coefficients $p_1(t, i), p_2(t, i), \dots, p_n(t, i)$ (that depend on F) such that $\sum_{j=1}^n p_j(t, i) = 1$ and*

$$P(t - T > x | T < t, X_{i:n} > t) = \sum_{j=1}^n p_j(t, i) P(t - X_{j:n} > x | X_{j:n} < t), \tag{2.5}$$

for all $0 \leq x \leq t$.

Proof. From lemma 2.1, we have

$$\begin{aligned} P(t - T > x | T < t, X_{i:n} > t) &= \frac{P(T < t - x, X_{i:n} > t)}{P(T < t, X_{i:n} > t)} \\ &= \frac{\sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k} \bar{F}^j(t) F^k(t-x)}{\sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k} \bar{F}^j(t) F^k(t)} \\ &= \sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} \frac{a_{j,k} \bar{F}^j(t) F^k(t)}{\sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k} \bar{F}^j(t) F^k(t)} \frac{F^k(t-x)}{F^k(t)} \\ &= \sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} \frac{a_{j,k} \bar{F}^j(t) F^k(t)}{\sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k} \bar{F}^j(t) F^k(t)} P(t - X_{k:k} > x | X_{k:k} < t) \\ &= \sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k}(t, i) P(t - X_{k:k} > x | X_{k:k} < t), \end{aligned} \tag{2.6}$$

where $P(t - X_{k:k} > x | X_{k:k} < t) = F^k(t-x)/F^k(t)$ and

$$a_{j,k}(t, i) = \frac{a_{j,k} \bar{F}^j(t) F^k(t)}{\sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k} \bar{F}^j(t) F^k(t)}.$$

Hence, if $a_k(t, i) = \sum_{j=n-i+1}^{n-k} a_{j,k}(t, i)$, from 2.6 we obtain

$$\begin{aligned} P(t - T > x | T < t, X_{i:n} > t) &= \sum_{k=1}^{i-1} a_k(t, i) P(t - X_{k:k} > x | X_{k:k} < t) \\ &= \sum_{k=1}^{i-1} a_k(t, i) \sum_{j=1}^n p_{j,k}(t) P(t - X_{j:n} > x | X_{j:n} < t), \\ &= \sum_{j=1}^n \left(\sum_{k=1}^{i-1} a_k(t, i) p_{j,k}(t) \right) P(t - X_{j:n} > x | X_{j:n} < t), \end{aligned}$$

such that $\sum_{j=1}^n p_{j,k}(t) = 1$ for all k . Upon denoting $p_j(t, i) = \sum_{k=1}^{i-1} a_k(t, i) p_{j,k}(t)$, we note that

$$\begin{aligned} \sum_{j=1}^n p_j(t, i) &= \sum_{j=1}^n \sum_{k=1}^{i-1} a_k(t, i) p_{j,k}(t) \\ &= \sum_{k=1}^{i-1} a_k(t, i) \left(\sum_{j=1}^n p_{j,k}(t) \right) \\ &= \sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k}(t, i) \\ &= \sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} \frac{a_{j,k} \bar{F}^j(t) F^k(t)}{\sum_{k=1}^{i-1} \sum_{j=n-i+1}^{n-k} a_{j,k} \bar{F}^j(t) F^k(t)} \\ &= 1. \end{aligned}$$

The vector of coefficients $\mathbf{p}(t, i) = (p_1(t, i), p_2(t, i), \dots, p_n(t, i))$ can be termed the conditional domination vector under the condition $(T < t, X_{i:n} > t)$. Note that the coefficients $p_j(t, i)$ need not all be nonnegative. The following examples show how these coefficients can be determined in specific cases. □

Example 2.3. *Let us consider the system $T = \max(X_1, \min(X_2, X_3))$, and let us assume that by time t the system has failed, but at least one component is still working, i.e., $\{T < t, X_{3:3} > t\}$. Then, for $0 \leq x \leq t$, we have*

$$\begin{aligned} P(T < t - x, X_{3:3} > t) &= P(X_1 < t - x, \min(X_2, X_3) < t - x, X_{3:3} > t) \\ &= P(X_1 < t - x, \min(X_2, X_3) < t - x, \max(X_1, X_2, X_3) > t) \\ &= P(X_1 < t - x, X_2 < t - x, X_3 > t) \end{aligned}$$

$$\begin{aligned}
&+ P(X_1 < t - x, X_2 > t, X_3 < t - x) \\
&= 2F^2(t - x)\bar{F}(t).
\end{aligned}$$

Therefore,

$$P(t - T > x | T < t, X_{3:3} > t) = \frac{F^2(t - x)}{F^2(t)} = P(t - X_{2:2} > x | X_{2:2} < t);$$

that is, $\mathbf{a}(t, \mathbf{3}) = (0, 1, 0)$. Hence, the system is equal in law to the past lifetime of a 2-component parallel system (an intuitive property). Finally, as the vector $\mathbf{p}_{2:2}(t)$ of coefficients in ?? of $X_{2:2}$ with $n = 3$ is $\mathbf{p}_{2:2}(t) = (0, \frac{3F(t) - 2F^2(t)}{6 - 3F(t)}, \frac{6 - 6F(t) + 2F^2(t)}{6 - 3F(t)})$, we have

$$P(t - T > x | T < t, X_{3:3} > t) = \frac{3F(t) - 2F^2(t)}{6 - 3F(t)} \bar{F}_{2:3}(x|t) + \frac{6 - 6F(t) + 2F^2(t)}{6 - 3F(t)} F(t) \bar{F}_{3:3}(x|t),$$

where $\bar{F}_{i:3}(x|t) = P(t - X_{i:3} > x | X_{i:3} < t)$, that is,

$$\mathbf{p}(t, \mathbf{3}) = \left(0, \frac{3F(t) - 2F^2(t)}{6 - 3F(t)}, \frac{6 - 6F(t) + 2F^2(t)}{6 - 3F(t)} \right).$$

References

- Goliforushani, S. and Asadi, M. (2011). Stochastic ordering among inactivity times of coherent systems, *Sankhya, Series B*, **73**, 241-262.
- Goliforushani, S., Asadi, M. and Balakrishnan, N. (2012). On the residual and inactivity times of the components of used coherent systems, *J. Appl. Probab.*, **49**, 385-404.
- Khaledi, B.E. and Shaked, M. (2007). Ordering conditional lifetimes of coherent systems, *J. Stat. Plann. Inf.*, **137**, 1173-1184.
- Kochar, S., Mukerjee, H. and Samaniego, F. J. (1999). The signature of a coherent system and its application to comparison among systems, *Naval Res. Logist.*, **46**, 507-523.
- Li, X. and Zhang, Z. (2008). Stochastic comparisons on general inactivity times and general residual life of k -out-of- n systems, *Commun. Statist. Simul. Comput.*, **37**, 1005-1019.
- Navarro, J., Balakrishnan, N. and Samaniego, F. J. (2008). Mixture representations of residual lifetimes of used systems, *J. Appl. Probab.*, **45**, 1097-1112.
- Navarro, J., Ruiz, J. M. and Sandoval, C. J. (2005). A note on comparisons among coherent systems with dependent components using signatures, *Statist. Probab. Lett.*, **72**, 179-185.

Samaniego, F. J. (1985). On closure of the IFR class under formation of coherent systems, *IEEE Trans. Reliab.*, **34**, 69-72.

Zhang, Z. (2010). Mixture representations of inactivity times of conditional coherent systems and their applications, *J. Appl. Probab.*, **47**, 876-885.



Stable Multiple Markov Processes

M. Taghipour¹, A.R. Nematollahi²

¹Department of Statistics, College of Sciences, Qom University

²Department of Statistics, College of Sciences, Shiraz University

Abstract: In this paper, we provide necessary conditions for a discrete-time symmetric α -stable processes to be linear 2-ple Markov. The aim of this paper is to extend the results given by Adler et al. (1990) to general multiple Markov processes, called linear multiple Markov processes. A necessary and sufficient condition based on the covariation for SaS processes to be linear multiple Markov is provided. A complete description of this class of covariation functions including the stationary case is given.

Keywords Markov stable processes , Time changed Levy motion, Covariation function.

Mathematics Subject Classification (2010): 60G10 60J05.

1 Introduction

The notion of multiple Markov property for stochastic processes has been studied by many authors; e.g. Hida (1960), Mandrekar (1974) and Kojo (1995). Let $X = \{X(t), t \in \mathbb{Z}\}$, where \mathbb{Z} stands for all integers, be a discrete-time zero mean process. In this paper, a linear multiple Markov is defined as follows.

Definition 1.1. *If the conditional expectation $E[X(t)|X(u), u \leq s]$ exists for all $s \leq t$ in \mathbb{Z} , then X is said to be “Linear Multiple Markov” (LMM) of order m , if*

$$E[X(t)|X(u), u \leq s] = \sum_{i=0}^{m-1} C_i(t, s)X(s-i), \quad (1.1)$$

for all $s \leq t$ in \mathbb{Z} , with probability one, where C_i are real numbers depending on t and s .

The class of symmetric α -stable (for short, SaS) processes, of which the Gaussian process is a special case, represents a natural generalization of the Gaussian processes, and provides a popular alternative for modeling of some popular time series which exhibit sharp spikes and outlying observations, see e.g. Nikias and Shao (1995). The random variable X is called symmetric α -stable, $0 < \alpha \leq 2$ if

$$E[e^{i\theta X}] = e^{-\mu|\theta|^\alpha}, \quad (1.2)$$

¹M. Taghipour : mtaghipour2002@gmail.com

for some $\mu > 0$. The stochastic processes $X = \{X(t), t \in R\}$ is S α S processes if and only if all linear combination are symmetric α -stable (i.e. for all $N \geq 1, a_i \in R, t_i \in T, \sum_{i=1}^N a_i X(t_i)$ is a S α S random variable). Each S α S process $X = \{X(t), t \in T\}$ has a representation of the form

$$X(t) = \int_U f_t(u) dZ(u), t \in T, \tag{1.3}$$

where (U, Σ, μ) is a measure space, and Z a random σ -additive set function on

$$\Sigma_\mu = \{E \in \Sigma : \mu(E) < \infty\}$$

which is independently scattered (i.e. $E \cap F = \phi \Rightarrow Z(E)$ and $Z(F)$ are independent) and S α S with $E [e^{i \theta Z(E)}] = e^{-\mu(E) |\theta|^\alpha}$. Furthermore, $f_t(u) \in L_\alpha(U, \Sigma, \mu)$, see Hardin Jr. (1982), Kanter (1972) and Kuelbs (1973). Let

$$u^{<\alpha>} = |u|^\alpha \operatorname{sgn} u.$$

Let $X(t)$ and $X(s)$ be jointly S α S with $1 < \alpha \leq 2$. The covariation function $X(t)$ on $X(s)$ is a real number

$$R(t, s) = \operatorname{Cov} [X(t), X(s)] = \int_U f_t(u) f_s(u)^{<\alpha-1>} d\mu(u), t \in T. \tag{1.4}$$

The covariation norm X is also defined by

$$\|X\|_\alpha = \{\operatorname{Cov}[X, X]\}^{\frac{1}{\alpha}}.$$

2 Main results

In this section, we provide a necessary and sufficient condition based on the covariation for S α S processes to be LMM. Then necessary conditions for discrete-time symmetric α -stable processes to be linear 2-ple Markov are also provided. We start with a lemma.

Lemma 2.1. *If X is LMM of order m , then*

$$E [X(t) | X(u), u \leq s] = E [X(t) | X(s), \dots, X(s - m - +1)], \quad s, t \in \mathbb{Z}.$$

Proof. Let $G_1 = \sigma\{X(s), X(s - 1), \dots, X(s - m - +1)\}$ and $G_2 = \sigma\{X(u), u \leq s\}$,

Then

$$E [X(t) | G_1] = E\{E[X(t) | G_2] | G_1\} = \sum_{i=0}^{m-1} C_i(t, s) X(s - i) = E[X(t) | G_2].$$

□

In the following theorem, we provide a necessary and sufficient condition for S α S processes to be LMM based on the covariation function, which extends the known result given by Adler et al. (1990, Theorem 2.1) to the linear multiple Markov processes.

Theorem 2.2. *X has LMM property of order m if and only if*

$$\text{Cov} \left[X(t) - \sum_{i=0}^{m-1} C_i(t, s) X(s-i), Y \right] = 0, \quad (2.1)$$

for all $Y \in \overline{SP} \{X(u), u \leq s\}, \forall s, t \in \mathbb{Z}, s \leq t$.

Proof. First note that X is orthogonal to X if and only if $\text{Cov}[X, Y] = 0$ (see Cambanis et al. (1988) Corollary 1.3). Let $L(S) = \overline{SP} \{X(u), u \in S\}$, where $S = \{u : u \leq s\}$, and $\tilde{X} = E[X(t) | X(u), u \leq s]$, then $L(S) \perp_p X(t) - \tilde{X}$, if and only if $\tilde{X} \in L(S)$ (see Cambanis et al. (1988) Proposition 1.5), which is true by (1.1). We recall that $X \perp_p Y$, when X is orthogonal to Y , $1 \leq p < \alpha$ for p^{th} order random variables variables (see Cambanis et al. (1988), p. 3). Therefore

$$L(S) \perp_p X(t) - \tilde{X} = X(t) - \sum_{i=0}^{m-1} C_i(t, s) X(s-i). \quad (2.2)$$

So $\text{Cov}[X(t) - \tilde{X}, Y] = 0$ for all $Y \in L(S)$, i.e.

$$\text{Cov} \left[X(t) - \sum_{i=0}^{m-1} C_i(t, s) X(s-i), Y \right] = 0$$

for all $Y \in \overline{SP} \{X(u), u \leq s\}, \forall s \leq t$, (see Cambanis et al. (1988), Corollary 1.3). \square

The coefficients $C_i(t, s)$ can be determined by the following simple procedure. Using Theorem 2.1, we substitutive Y by the values $X(s-i), i = 0, 1, \dots, m-1$, to reduce

$$\text{Cov} \left[X(t) - \sum_{i=0}^{m-1} C_i(t, s) X(s-j), Y \right] = 0, \quad j = 0, 1, \dots, m-1.$$

Since covariation is linear in the first argument, we have

$$\text{Cov}[X(t), X(s-j)] = \sum_{i=0}^{m-1} C_i(t, s) \text{Cov}[X(s-i), X(s-j)], \quad j = 0, 1, \dots, m-1.$$

So $R(t, s - j) = \sum_{i=0}^{m-1} C_i(t, s) R(s - i, s - j)$, $j = 0, 1, \dots, m - 1$, which can be written in the following matrix form $RC = r$, where

$$R = \begin{bmatrix} R(s, s) & R(s - 1, s) & \dots & R(s - m + 1, s) \\ R(s, s - 1) & R(s - 1, s - 1) & \dots & R(s - m + 1, s) \\ \vdots & \vdots & \ddots & \vdots \\ R(s, s - m + 1) & R(s - 1, s - m + 1) & \dots & R(s - m + 1, s - m + 1) \end{bmatrix},$$

$$C = \begin{bmatrix} C_0(t, s) & C_1(t, s) & \dots & C_{m-1}(t, s) \end{bmatrix}^t, r = \begin{bmatrix} R(t, s) & R(t, s - 1) & \dots & R(t, s - m + 1) \end{bmatrix}^t.$$

For example, let $m = 1$ (linear 1-ple Markov), we have

$$E[X(t)|X(u), u \leq s] = C_0(t, s) X(s),$$

where $C_0(t, s)$ is obtained by solving the following equation $R(t, s) = R(s, s) C_0(t, s)$. Hence X has LMM if and only if

$$Cov[X(t) - \frac{R(t, s)}{R(s, s)} X(s), Y] = 0 \quad \forall s \leq t, Y \in \overline{SP}\{X(u), u \leq s\}.$$

For $m = 2$ (linear 2-ple Markov), we have

$$E[X(t)|X(u), u \leq s] = C_0(t, s) X(s) + C_1(t, s) X(s - 1),$$

where $C_0(t, s)$ and $C_1(t, s)$ will be obtained by solving the following equation

$$\begin{vmatrix} R(s, s) & R(s - 1, s) & C_0(t, s) \\ R(s, s - 1) & R(s - 1, s - 1) & C_1(t, s) \end{vmatrix} = \begin{vmatrix} R(t, s) \\ R(t, s - 1) \end{vmatrix}$$

It yields

$$\begin{cases} C_0(t, s) = \frac{R(s-1, s-1) R(t, s) - R(s-1, s) R(t, s-1)}{R(s, s) R(s-1, s-1) - R(s-1, s) R(s, s-1)} \\ C_1(t, s) = \frac{R(s, s) R(t, s-1) - R(s, s-1) R(t, s)}{R(s, s) R(s-1, s-1) - R(s-1, s) R(s, s-1)} \end{cases} \quad (2.3)$$

Remark 2.3. From Miller (1978, Corollary 3.2), and Lemma 2.1, we conclude that

$$E[X(t)|X(s), \dots, X(s - m + 1)] = E[X(t)|X(u), u \leq s] = \sum_{i=0}^{m-1} C_i(t, s) X(s - i),$$

if and only if, for all r_1, \dots, r_m

$$\int (x(t) - C_0(t, s)x(s) - \dots - C_{m-1}(t, s)x(s - m + 1) (r_1 x(s) - \dots - r_m x(s - m + 1)))^{\alpha-1} \mu(dx) = 0.$$

Setting $r_1 = 1, r_2 = \dots = r_m = 0$, it reduces to

$$\int x(t)x(s)^{\alpha-1} \mu(dx) = C_0(t, s) \int x(s) x(s)^{\alpha-1} \mu(dx) + \dots + C_{m-1}(t, s) \int x(s - m + 1)^{\alpha-1} x(s) \mu(dx),$$

which means

$$R(t, s) = C_0(t, s) R(s, s) + C_1(t, s) R(s, s - 1) + \dots + C_{m-1}(t, s) R(s - m + 1, s).$$

Similarly, let $r_2 = 1, r_1 = r_3 \dots = r_m = 0$, then

$$R(t, s - 1) = C_0(t, s)R(s, s - 1) + C_1(t, s) R(s - 1, s - 1) + \dots + C_{m-1}(t, s) R(s - m + 1, s - 1),$$

Continuing this process, let $r_m = 1, r_1 = r_2 \dots = r_{m-1} = 0$, then

$$R(t, s-m+1) = C_0(t, s)R(s, s-m+1) + C_1(t, s) R(s-1, s-m+1) + \dots + C_{m-1}(t, s) R(s-m+1, s-m+1),$$

which we have already obtained from another approach. For the sake of simplicity, we now consider only the case of order 2 ($m=2$). Extension to the general case of m , will be followed by a straightforward similar procedure.

Corollary 2.4. For $m = 2$, (linear 2-ple Markov), we have

$$\begin{vmatrix} R(t_2 - 1, t_2) & R(t_2, t_2) & R(t_3, t_2) \\ R(t_2 - 1, t_2 - 1) & R(t_2, t_2 - 1) & R(t_3, t_2 - 1) \\ R(t_2 - 1, t_1) & R(t_2, t_1) & R(t_3, t_1) \end{vmatrix} = 0, \quad \forall t_1 \leq t_2 \leq t_3. \tag{2.4}$$

Proof. Take $Y = X(u)$, in theorem 2.1 and substituting $C_0(t, s)$ and $C_1(t, s)$ by (2.3), the proof is complete. □

Theorem 2.5. If X is linear 2-ple Markov $S\alpha S$ process, then the covariation function R with $R(t, s) \neq 0$ for all $s < t$, is of the form

$$R(t, s) = H_1(t) K_1(s)^{\langle \alpha-1 \rangle} + H_2(t) K_2(s)^{\langle \alpha-1 \rangle}, \quad \forall s, t \in \mathbb{Z}, s \leq t, \tag{2.5}$$

where the functions H_i and K_i are unique up to a multiplicative constant, and satisfy in the following conditions

- i) $H_1(t) K_1(t)^{\langle \alpha-1 \rangle} + H_2(t) K_2(t)^{\langle \alpha-1 \rangle} > 0$
- ii) $|H_1(t) K_1(s)^{\langle \alpha-1 \rangle} + H_2(t) K_2(s)^{\langle \alpha-1 \rangle}| \leq [H_1(t) K_1(t)^{\langle \alpha-1 \rangle} + H_2(t) K_2(t)^{\langle \alpha-1 \rangle}]^{\frac{1}{\alpha}} [H_1(s) K_1(s)^{\langle \alpha-1 \rangle} + H_2(s) K_2(s)^{\langle \alpha-1 \rangle}]^{1-\frac{1}{\alpha}}$

Proof. Let $t_0, s_0 (s_0 \leq t_0)$ be any points of \mathbb{Z} . Consider the following scenario

- i) $s < t \leq s_0 < t_0$ ii) $s < s_0 \leq t < t_0$ iii) $s < s_0 \leq t_0 < t$
- iv) $s_0 < t_0 \leq s < t$ v) $s_0 < s < t_0 \leq t$ vi) $s_0 < s < t < t_0$

We only consider the case (i), the proof of the other cases is similar. Let $s < t \leq s_0 < t_0$, then now (2.5) holds by taking

$$H_1(t) = \frac{R(s_0, s_0) R(t, t) - R(t, s_0) R(s_0, t)}{R(s_0, s_0) R(t_0, t) - R(t_0, s_0) R(s_0, t)},$$

$$H_2(t) = \frac{R(t, s_0) R(t_0, t) - R(t_0, s_0) R(t, t)}{R(s_0, s_0) R(t_0, t) - R(t_0, s_0) R(s_0, t)},$$

$$K_1(s)^{<\alpha-1>} = R(t_0, s), \quad K_2(s)^{<\alpha-1>} = R(s_0, s).$$

The proof is now complete by noting that $0 < R(t, t)$ and (ii) holds by using the well-known Holder inequality. Now it remains to show that (2.5) is a covariation function. The simplest way of showing this is to make a SaS process with covariation function (2.5). We recall that the SaS Levy motion $L = \{L(t), t \geq 0\}$ is a process with stationary independent increments, $L(0) = 0$, a.s., and for all real r and $t, s \geq 0$,

$$E[\exp\{ir(L(t) - L(s))\}] = \exp\{-|r|^\alpha |t - s|\}.$$

Also if $\tau(t)$ is positive and non-decreasing on T , and $H(t)$ is positive on T , then the time change of SaS Levy motion

$$X(t) = H(t)L(\tau(t)), t \in T,$$

is Markov and for $s < t$, the conditional distribution of $X(t)$ given $X(s)$ is α -stable and symmetric, See Adler et al. (1990). Now consider the time changes $\tau_i(t) = \left[\frac{K_i(t)}{H_i(t)}\right]^{\alpha-1}$, $i = 1, 2$, where $\frac{K_i(t)}{H_i(t)}$ is positive and nondecreasing on T . Now let $X(t) = X_1(t) + X_2(t)$, where $X_i(t) = H_i(t)L_i(\tau_i(t))$, $i = 1, 2$, are two independent time change Levy motions. Then for all $s < t$, we have

$$\begin{aligned} R(t, s) &= Cov[X_1(t) + X_2(t), X_1(s) + X_2(s)] \\ &= H_1(t) H_1(s)^{<\alpha-1>} \min(\tau_1(t), \tau_1(s)) + H_2(t) H_2(s)^{<\alpha-1>} \min(\tau_2(t), \tau_2(s)). \end{aligned}$$

Since $\tau_i(t)$, $i = 1, 2$ is nondecreasing, positive and both $K_i(t)$ and $H_i(t)$ have the same sign, then

$$R(t, s) = H_1(t) K_1(s)^{<\alpha-1>} + H_2(t) K_2(s)^{<\alpha-1>}.$$

□

Remark 2.6. Note that the sum of two independent time change Levy motions $X(t) = X_1(t) + X_2(t)$ is linear 2-ple Markov, as we have

$$X(t) = \frac{H_1(t)}{H_1(s)} X_1(s) + H_1(t) [L_1(\tau_1(t)) - L_1(\tau_1(s))] + \frac{H_2(t)}{H_2(s-1)} X_2(s-1) [L_2(\tau_2(t)) - L_2(\tau_2(s-1))].$$

The conditional distribution of the $X(t)$ given $\sigma\{X(u), u \leq s\}$ depends only on the $\sigma\{X(s), X(s-1)\}$, and so

$$E \exp\{ir X(t) | X(u), u \leq s\} = E \exp\{ir X(t) | X(s), X(s-1)\}.$$

Remark 2.7. When X is stationary, then $R(t, s)$ depends only on $(t - s)$ (i.e., $R(t, s) = R(t - s)$ for all $t, s \in \mathbb{R}$) and we have the covariation stationary terminology. In the special case, if X is linear 2-ple Markov stationary $S\alpha S$ process, then by Corollary 2.1,

$$\begin{vmatrix} R(-1) & R(0) & R(t_3 - t_2) \\ R(0) & R(1) & R(t_3 - t_2 + 1) \\ R(t_2 - t_1 - 1) & R(t_2 - t_1) & R(t_3 - t_1) \end{vmatrix} = 0.$$

Take $t = t_3 - t_2$ and $s = t_2 - t_1$ and by taking $s = 2$, we get the following a second-order linear homogeneous recurrence relation $R(t+2) = aR(t+1) + bR(t)$, $t > 2$, where $a = \frac{R(2)R(-1) - R(1)R(0)}{R(-1)R(1) - R(0)^2}$, $b = \frac{R(1^2) - R(2)R(0)}{R(-1)R(1) - R(0)^2}$.

References

- Adler, R.J., Cambanis, S. and Samorodnitsky, G. (1990). On stable Markov processes, *Stochastic Process. Appl.* 34 1-17., *Statistical Models and Methods for Biomedical and Technical Systems*, Birkhuser, Boston, pp. 319-330.
- Cambanis, S., Hardin, Jr. and Weron, A. (1988). Innovations and Wold decompositions of stable sequences, *Probab. Theory Rel. Fields* 79, 1-27.
- Hardin, Jr. (1982). On the spectral representation of symmetric stable processes, *J. Multivariate Anal.*, **12**, 385-401.
- Hida, T. (1960). Conational representation of Gaussian processes and their applications, *Mem. Coll. Sci. Univ. Kyoto*, **33**, 109-155.
- Kanter, M. (1972). Linear sample spaces and stable processes, *J. Funct. Anal.*, **9**, 441-456.
- Kojo, K. (1995). On the notion of multiple Markov $S\alpha S$ processes, *Hiroshima Mathematical Journal*, **25**, 143-157.

-
- Kuelbs, J. (1973). A representation theorem for symmetric stable processes and stable measures on H , *Z. Wahrsch. Verw. Gebiete*, **26**, 259-271.
- Mandrekar, V. (1974). On the multiple Markov property of the LEVY-HIDA for Gaussian processes, *Nagoya Math. J*, **54**, 69-78.
- Miller, G. (1978). Properties of certain symmetric stable distributions, *J. Multivariate Anal*, **8**, 346-360.
- Nikias, C.L. and Shao, M. (1995). *Signal Processing with Alpha-Stable Distributions and Applications*, New York: Wiley.



Weighted Integration to Solutions of SDE's with uniformly elliptic diffusion

M. Tahmasebi ¹

¹Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, P.O. Box 14115-134, Tehran, Iran

Abstract: In this work we prove an integration by parts formula for the solution of the SDE monotone drifts in the Wiener space. We construct an approximating sequence of SDE's with globally Lipschitz drifts and obtain a uniform bound for the integral of their solutions.

Keywords stochastic differential equation, monotone drift, Malliavin calculus.

Mathematics Subject Classification (2010): 60H10.

1 Introduction

Assuming the nondegeneracy condition one can derive some integration by parts formula on the Wiener space (see e.g. [Nualart \(2006\)](#)). This formula has many applications for example in financial mathematics. It is often of interest to investors to derive an option pricing formula and to know its sensitivity with respect to various parameters. The integration by parts formula obtained from Malliavin calculus can transform the derivative of the option price into weighed integral of random variables. This gives much more accurate and fast converging numerical solution estimates than obtained by the classical methods [Kohatsu-Higa and Montero \(2004\)](#); [Bavouzet and \(2006\)](#). The interested reader could see [Alòs and Ewald \(2008\)](#); [Alò, León and Vives \(2007\)](#); [Talay \(2002\)](#); [Bally and Talay \(1996\)](#).

The SDE we consider has not global Lipschitz coefficients. Such equations mostly come from finance and biology and also dynamical systems and are more challenging when considered on infinite dimensional spaces. (see e.g. [Bahlali \(1999\)](#); [Zangeneh \(1995\)](#); [Gyöngy and Millet \(2007\)](#))

In this paper, we consider an SDE with locally Lipschitz coefficients and uniformly elliptic diffusion. In [Tahmasebi and Zamani \(2013\)](#), we have proved the uniqueness and existence of a solution X_t to this equation. Since the drift of the SDE is not globally Lipschitz, we will construct a sequence of SDEs with globally Lipschitz drifts whose solutions have uniformly bounded Malliavin derivatives with respect to n , and converge to X_t almost everywhere. In this way we can apply the classical Malliavin calculus to these solutions, and by the uniform boundedness of the moments of inverses of Malliavin

¹Tahmasebi: tahmasebi@modares.ac.ir

covariance matrices and the convergence result we are able to prove an integration by parts formula in the Wiener space.

The organization of the paper is as follows. In section 2, we present some notions of Malliavin calculus, as stated in [Nualart \(2006\)](#), and by use of them we formulate our main results. In section 3, we will prove the integration by parts formula in the Wiener space.

2 Formulation of main results

Let Ω denote the Wiener space $C_0([0, T]; R^d)$. We furnish Ω with the $\| \cdot \|_\infty$ -norm making it a (separable) Banach space. Consider (Ω, \mathcal{F}, P) a complete probability space, in which \mathcal{F} is generated by the open sets of the Banach space, W_t is a d -dimensional Brownian motion, and \mathcal{F}_t is the filtration generated by W_t .

By $H := L^2([0, T];^d)$ we denote a Hilbert space. For $k, p \geq 1$, denote by k,p the domain of the k th order Malliavin derivative operator with respect to the norm

$$\| F \|_{k,p} = \left[E|F|^p + \| D^{i_1, \dots, i_k} F \|_{L^p(\Omega; H^{\otimes k})}^p \right]^{1/p},$$

and define ${}^\infty := \bigcap_{k,p} {}^{k,p}$.

Now consider the following stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0. \tag{2.1}$$

where $b :^d \rightarrow^d$ and $\sigma :^d \rightarrow M_{d \times d}()$ are measurable functions. The function σ is C^∞ and all of its derivatives of order greater or equal to one are bounded. The function σ is globally Lipschitz with Lipschitz constant k_1 . $\sigma\sigma^*$ is uniformly continuous with modulus of continuity $\theta(\cdot)$. Also for some positive constant C_2

$$|\sigma^*(x)u|^2 \leq C_2|u|^2 \tag{2.2}$$

Note that in this paper, the function b is not considered globally Lipschitz.

In ([Nualart , 2006](#), Theorem 2.2.2 and Corollary 2.2.1) Nualart shows that SDEs which have globally Lipschitz coefficients with polynomial growth for all of their derivatives, have strong unique solutions in ${}^\infty$. He derives the linear equation which the Malliavin derivative satisfy as well. In [?Tahmasebi and Zamani \(2013\)](#) assuming some conditions for the nonglobally Lipschitz coefficients, we show the existence of a unique strong solution X_t in ${}^\infty$ to SDE (2.1). In this paper, we are not concerned with the existence and uniqueness result and just assume that the SDE (2.1) has a unique strong solution in ${}^\infty$. We denote by Δ the second-order differential operator associated to SDE (2.1):

$$= \frac{1}{2} \sum_{i,j=1}^d (\sigma\sigma^*)^i_j(x) \partial_i \partial_j + \sum_{i=1}^d b^i(x) \partial_i$$

Consider the following stochastic differential equations

$$dX_t^n = b_n(X_t^n)dt + \sigma(X_t^n)dW_t, \quad X_0^n = x_0. \tag{2.3}$$

where the functions b_n are globally Lipschitz and all of their derivatives have polynomial growth; i.e., for each $x \in^d$ and each multi-index α with $|\alpha| = m$, there exist positive constants q_m and Γ_m which are independent of n and

$$|\partial_\alpha b_n(x)|^2 \leq \Gamma_m(1 + |x|^{q_m}) \tag{2.4}$$

As we pointed out, by Nualart (2006) there exist unique strong solutions $X_t^n \in^\infty$ for SDEs (2.3). Also, for $r \leq t$,

$$= \sigma^i(X_r^n) + \int_r^t \nabla b_n^i(X_s^n).ds + \int_r^t \nabla \sigma_l^i(X_s^n).dW_s^l \tag{2.5}$$

and for $r > t, = 0$. Here $u.C$ denotes the product C^*u for a vector u and a matrix C , for example $\nabla f(x) = \sum_{l=1}^d \nabla_l f(x)$. We used the upper index to show a specific row, and the subindex to show a specific column of a matrix. As before denote the infinitesimal operator associated to these SDE's by .

Throughout the paper we assume the following Hypothesis.

Hypothesis 2.1. *For each $p \geq 1$, there exists a positive constant c_p such that the sequence $\{X_t^n\}_{n \geq 1}$ converges to X_t in $L^p(\Omega)$ and*

$$\sup_{n \geq 1} \sup_{0 \leq t \leq T} \left[\|\|_H^p + |X_t^n|^p \right] \leq c_p. \tag{2.6}$$

For every $0 \leq t \leq T$, denote the Malliavin covariance matrix of X_t^n and X_t by $Q_n(t)$ and $Q(t)$, respectively. Nualart (Nualart , 2006, Theorem 2.3.3 and its proof) have shown that if that, the Hörmander condition holds, the solutions X_t^n have a.s. invertible Malliavin covariance matrices, infinitely differentiable densities, and the nondegeneracy condition holds for them.

3 Integration by parts formula

In this section, we prove Theorem 3.1. By the integration by parts formula in (Nualart , 2006, Proposition 2.1.4.) and Hypothesis ??, there exists a family of random variables $\{L_\beta\}$ depending on multiindices β of length strictly larger than 1 with coordinates $\beta_j \in \{1, \dots, d\}$, such that for every $G \in^\infty$

$$[\partial_\alpha g(X_t^n)G] = E[g(X_t^n)L_\alpha^n(X_t^n, G)], \tag{3.1}$$

and

$$\| L_\alpha^n(X_t^n, G) \|_p \leq c_{p,q} \| \det((\cdot)^{-1}) \|_{\beta_0}^m \| DX_t^n \|_{k,\gamma}^n \| G \|_{k,q} \tag{3.2}$$

where

$$L_{(i)}^n(X_t^n, G) = \sum_{j=1}^m \delta \left(G^{(-1)}_j^i \right),$$

and δ denotes the adjoint of the Malliavin derivative operator D .

Theorem 3.1. *Let $g \in C^{m+1}$, and all of its derivatives be bounded and have polynomial growth. If Hypotheses 2.1 hold and for every $p \geq 1$, $\det(\cdot)^{-p}$ is uniformly bounded, then for every $G \in \infty$ and every multiindex $\alpha = (\alpha_1, \dots, \alpha_d)$ with $|\alpha| = m$, there exists a function $H_\alpha(X_t, G)$ such that*

$$\left(\partial_\alpha g(X_t) G \right) = \left(g(X_t) H_\alpha(X_t, G) \right), \quad (3.3)$$

where $H_\alpha = H_{\alpha_k}(H_{(\alpha_1, \dots, \alpha_{k-1})})$.

References

- Nualart, D. (2006) *The Malliavin Calculus and Related Topics*, second ed., Springer Verlag, Berlin.
- Hayashi, M., Kohatsu-Higa, A. and Yûki, G.(2012) Local Hölder continuity property of the densities of solutions of SDEs with singular coefficients, To appear in *Journal of Theoretical Probability*, Submitted on 6 Jun 2012. <http://arxiv.org/abs/1206.1117v1> .
- Kohatsu-Higa, A. and Montero, M., (2004) *Malliavin Calculus in Finance*, Handbook of computational and numerical methods in finance, Birkhäuser Boston, Boston, MA, pp. 111–174.
- Bavouzet, M.P. and Messaoud, M., (2006) Computation of Greeks using Malliavin's calculus in jump type market models, *Electronic Journal of Probability* **11(10)**, pp. 276–300.
- Alòs, E. and Ewald, C.O.,(2008) Malliavin differentiability of the Heston volatility and applications to option pricing, *Adv. in Appl. Probab.* **40(1)**, pp. 14–162.
- Alòs, E., León, J.A., and Vives, J., (2007) On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility, *Finance Stoch.* **11(4)**, pp. 571–589.
- M. Tahmasebi and S. Zamani, (2013) Weak differentiability of solutions to SDEs with semi-monotone drifts, preprint in Bull. Iranian Math. Soc.
- Talay, D. (2002) Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme, *Markov Processes relat. Fields* **8** pp. 1–36.

-
- Bally, V., and D. Talay, (1996) The law of the Euler scheme for stochastic differential equations(I): convergence rate of the distribution function, *Probability Theory and Related Fields* **104(1)**, pp. 43–60.
- Bahlali, K. (1999) Flows of homeomorphisms of stochastic differential equations with measurable drifts, *Stochastics and Stochastics Reports* **67** pp. 53–82.
- Zangeneh, B. Z. (1995). Semilinear stochastic evolution equations with monotone nonlinearities, *Stochastics and Stochastics Reports* **3** pp. 129–174.
- Gyöngy, I. and Millet, A. (2007). Rate of convergence of implicit approximations for stochastic evolution equations, *Stochastic Differential Equations: Theory and Applications, A volume in honor of professor Boris L. Rosovskii, Interdisciplinary Mathematical Sciences, 2*, World Scientific pp. 281–310.



Information Measures for Weighted Distributions of Exp-G Family

Saeid Tahmasebi¹, Ali Akbar Jafari²

¹Department of Statistics, Persian Gulf University, Bushehr, Iran

²Department of Statistics, Yazd University, Iran

Abstract: Weighted distributions are widely used in statistical sciences and arise when the observations generated from a stochastic process are recorded with some weight function. In this article, we derive analytical expressions of β -entropy for different weighted distributions of exp-G family. Further, we present general forms of the residual entropy measures of order β for this family.

Keywords Exp-G family, Tsallis entropy, Weighted distributions.

Mathematics Subject Classification (2010): 60E05 94A17.

1 Introduction

The concept of weighted distribution introduced by Rao (1965) is widely used in statistical sciences. Weighted distributions arise when the observations generated from a stochastic process are recorded with some weight function. Also, these distributions have seen much use as a tool in the selection of appropriate models for observed data drawn without a proper frame. Some examples of generating weighted distributions and their applications is given in Patil (2002).

To introduce the concept of a weighted distribution, using a weight function, $w(x) \geq 0$, suppose X is a non-negative random variable with probability density function (pdf) $f_{\theta}(x)$. The random variable X_w is called the weighted version of X iff it has the following pdf:

$$f_{\theta}^w(x) = \frac{w(x)f_{\theta}(x)}{E_{\theta}[w(X)]}, \quad 0 \leq x < \infty, \quad (1.1)$$

where $0 < E_{\theta}[w(X)] < \infty$.

The truncated exponential distribution in the interval $[0, 1]$ with parameter λ is given by

$$F_{\lambda}^*(x) = \frac{1 - \exp(-\lambda x)}{1 - \exp(-\lambda)}, \quad \lambda > 0; \quad x \in [0, 1]. \quad (1.2)$$

¹Saeid Tahmasebi: tahmasebi@pgu.ac.ir

A new class of distributions is defined by [Barreto-Souza and Simas \(2010\)](#) as follows. Let $G(x; \theta)$ be the cumulative distribution function (cdf) of a continuous random variable with parameter θ , then the exp-G distribution is given by

$$F_{\lambda}^G(x) = F_{\lambda}^*(G(x; \theta)). \quad (1.3)$$

Therefore, the pdf of this distribution is

$$f_{\lambda, \theta}(x) = \frac{\lambda}{1 - e^{-\lambda}} g(x; \theta) \exp\{-\lambda G(x; \theta)\}, \quad x \in [0, 1], \quad (1.4)$$

where $g(x; \theta) = \frac{\partial G(x; \theta)}{\partial x}$.

If X is a none-negative continuous random variable with probability density function (pdf) $f(x)$, then Shannon's entropy of X is defined as

$$H(f) = E[-\log(f(X))] = - \int_0^{+\infty} f(x) \log(f(x)) dx,$$

and this is usually referred to as the continuous entropy (or differential entropy).

The concept of entropy originated in the nineteenth century by [Shannon \(1948\)](#) and many extensions of Shannon's entropy have been studied by various authors. [Renyi \(1961\)](#), [Havrda and Charvat \(1967\)](#), [Tsallis \(1988\)](#), [Cover and Thomas \(1991\)](#), [Kapur \(1994\)](#), and [Nadarajah and Zografos \(2003\)](#) are among the researchers in this area. β -entropy was first proposed by [Havrda and Charvat \(1967\)](#) in the context of cybernetics theory, and later applied in physical problems by [Tsallis \(1988\)](#). Hence β -entropy is also known as Tsallis entropy and for a none-negative continuous random variable X with pdf $f(x)$ is defined as

$$H_{\beta}(f) = \frac{1}{\beta - 1} (1 - \int_0^{+\infty} [f(x)]^{\beta} dx), \quad \beta > 0, \quad \beta \neq 1, \quad (1.5)$$

which is a generalization for the Shannon's entropy, since

$$H(f) = \lim_{\beta \rightarrow 1} H_{\beta}(f).$$

The Shannon entropy for exp-G family is obtained by [Barreto-Souza and Simas \(2010\)](#) as

$$H(f) = 1 - \frac{\lambda}{e^{\lambda} - 1} - \log\left[\frac{\lambda}{1 - e^{-\lambda}}\right] - E[\log g(G^{-1}(W))],$$

where W follows truncated exponential distribution with parameter λ . Recently, [Riabi et al. \(2010\)](#) obtained some results related to the β -entropy for different weighted versions of Pareto distribution. In this paper, we consider a random variable, X with pdf in (1.3) and we will propose different weighted distributions of exp-G family using different weight function, $w(x)$. Also we derive the β -entropy of these weighted distributions in Section 2. Further, in Section 3, we provide some results of residual entropy measures of order β for exp-G family.

2 β -entropy for a weighted exp-G family

- Let $w_1(x) = (F(x))^{a-1}(1 - F(x))^{b-1}$ as a weight function; then from (1.1) we have

$$\begin{aligned} f_{\lambda,\theta}^{w_1}(x) &= \frac{(F(x))^{a-1}(1 - F(x))^{b-1}f(x)}{E[(F(X))^{a-1}(1 - F(X))^{b-1}]} \\ &= \frac{\lambda g(x)[1 - \exp(-\lambda G(x))]^{a-1} \exp(-\lambda G(x))[\exp(-\lambda G(x)) - \exp(-\lambda)]^{b-1}}{B(a, b)(1 - \exp(-\lambda))^{a+b-1}}, \end{aligned}$$

where $B(a, b)$ is the beta function. This is the beta exp-G family and assuming that the inverse $G^{-1}(x)$ exists, its β -entropy is obtained as

$$\begin{aligned} H_\beta(f^{w_1}) &= \frac{1}{\beta - 1} - \frac{\lambda^{\beta-1}B(\beta(a - 1) + 1, \beta(b - 1) + 1)}{(1 - e^{-\lambda})^{\beta-1}(\beta - 1)[B(a, b)]^\beta} \times \\ &\quad E[g(G^{-1}(\frac{-\log(1 - T_1(1 - e^{-\lambda}))}{\lambda}))](1 - T_1(1 - e^{-\lambda}))]^{\beta-1}, \end{aligned} \tag{2.1}$$

where T_1 has the beta distribution with parameters $\beta(a - 1) + 1$ and $\beta(b - 1) + 1$. Further

$$\begin{aligned} H(f^{w_1}) &= \lim_{\beta \rightarrow 1} H_\beta(f^{w_1}) = -\log[\frac{\lambda}{(1 - e^{-\lambda})B(a, b)}] + (a - 1)[\psi(a + b) - \psi(a)] \\ &\quad + (b - 1)[\psi(a + b) - \psi(b)] - E[\log g(G^{-1}(\frac{-\log(1 - T(1 - e^{-\lambda}))}{\lambda})))] \\ &\quad + E[\log(1 - T(1 - e^{-\lambda}))], \end{aligned}$$

where ψ is the digamma function and T follows beta distribution with parameter a and b .

Remark 2.1. If $a = 1$ and $b = 1$, then from (2.1) the β -entropy for exp-G distribution is given by

$$H_\beta(f^{w_1}) = \frac{1}{\beta - 1} \{1 - [\frac{\lambda}{1 - e^{-\lambda}}]^{\beta-1} E[g(G^{-1}(W))e^{-\lambda W}]^{\beta-1}\}.$$

where W follows truncated exponential distribution with parameter λ and cdf is given in (1.2).

- Consider the weighted function as $w_2(x) = (-\log F(x))^{a-1}$, from (1.1) we obtain the pdf as

$$f_{\lambda,\theta}^{w_2}(x) = \frac{[-\log F(x)]^{a-1}f(x)}{E_{f(x)}[(-\log F(X))^{a-1}]} = \frac{[-\log(1 - \exp(-\lambda G(x)))]^{a-1} \lambda g(x) \exp(-\lambda G(x))}{\Gamma(a)(1 - e^{-\lambda})}.$$

Hence,

$$H_\beta(f^{w_2}) = - \frac{\lambda^{\beta-1}\Gamma(\beta(a - 1) + 1)}{(\beta - 1)[(1 - e^{-\lambda})\Gamma(a)]^\beta} E[g(G^{-1}(\frac{-\log(1 - \exp(-T_2))}{\lambda}))](1 - \exp(-T_2))]^{\beta-1}$$

$$+ \frac{1}{\beta - 1}, \quad (2.2)$$

where T_2 has the gamma distribution with parameters $\beta(a - 1) + 1$ and 1.

- For $w_3(x) = (-\log(1 - F(x)))^{b-1}$, from (1.1) we have a weighted density function as

$$f_{\lambda, \theta}^{w_3}(x) = \frac{[-\log(1 - F(x))]^{b-1} f(x)}{E_{f(x)}[(-\log(1 - F(X)))^{b-1}]} = \frac{[-\log(1 - \frac{1-e^{-\lambda G(x)}}{1-e^{-\lambda}})]^{b-1} \lambda g(x) \exp(-\lambda G(x))}{\Gamma(b)(1 - e^{-\lambda})}.$$

Also,

$$H_{\beta}(f^{w_3}) = \frac{1}{\beta - 1} - \frac{(\frac{\lambda}{1-e^{-\lambda}})^{\beta-1} \Gamma(\beta(b-1) + 1)}{(\beta - 1)[\Gamma(b)]^{\beta}} \times \quad (2.3)$$

$$E[g(G^{-1}(\frac{-\log[1 - (1 - e^{-\lambda})(1 - \exp(-T_3))]}{\lambda})) (1 - (1 - e^{-\lambda})(1 - \exp(-T_3)))]^{\beta-1},$$

where T_3 has the gamma density with parameters $\beta(b - 1) + 1$ and 1.

- Let $w_4(x) = (F(x))^a$ as a weight function, then from (1.1) we have

$$f_{\lambda, \theta}^{w_4}(x) = \frac{[F(x)]^a f(x)}{E_{f(x)}[(F(X))^a]} = \frac{[1 - \exp(-\lambda G(x))]^a \lambda g(x) (a + 1) \exp(-\lambda G(x))}{(1 - e^{-\lambda})^{a+1}}.$$

Further,

$$H_{\beta}(f^{w_4}) = \frac{1}{\beta - 1} - \frac{\lambda^{\beta-1} [a + 1]^{\beta} E[g(G^{-1}(\frac{-\log[1 - (1 - e^{-\lambda})T_4]}{\lambda})) (1 - (1 - e^{-\lambda})T_4)]^{\beta-1}}{(1 - e^{-\lambda})^{\beta-1} (\beta - 1)(a\beta + 1)}, \quad (2.4)$$

where T_4 has the beta distribution with parameters $a\beta + 1$ and 1.

- For $w_5(x) = (1 - F(x))^b$, we have a weighted density function as

$$f_{\lambda, \theta}^{w_5}(x) = \frac{[1 - F(x)]^b f(x)}{E_{f(x)}[(1 - F(X))^b]} = \frac{[1 - \frac{1 - \exp(-\lambda G(x))}{1 - e^{-\lambda}}]^b \lambda g(x) \exp(-\lambda G(x)) (b + 1)}{1 - e^{-\lambda}}.$$

Hence,

$$H_{\beta}(f^{w_5}) = \frac{1}{\beta - 1} - \frac{\lambda^{\beta-1} [b + 1]^{\beta} E[g(G^{-1}(\frac{-\log[1 - (1 - e^{-\lambda})T_5]}{\lambda})) (1 - (1 - e^{-\lambda})T_5)]^{\beta-1}}{(1 - e^{-\lambda})^{\beta-1} (\beta - 1)(b\beta + 1)}, \quad (2.5)$$

where T_5 has the beta distribution with parameters 1 and $b\beta + 1$

- Suppose that the weight function is $w_6(x) = [(1 - (F(x))^c]^r$; then from (1.1) we have

$$f_{\lambda,\theta}^{w_6}(x) = \frac{[1 - (F(x))^c]^r f(x)}{E_{f(x)}[(1 - (F(X))^c)^r]} = \frac{c[1 - (\frac{1 - \exp(-\lambda G(x))}{1 - e^{-\lambda}})^c]^r \lambda g(x) \exp(-\lambda G(x))}{B(\frac{1}{c}, r + 1)(1 - e^{-\lambda})}.$$

Further,

$$H_\beta(f^{w_6}) = \frac{1}{\beta - 1} - \frac{(c\lambda)^{\beta-1} B(\frac{1}{c}, r\beta + 1) E[g(G^{-1}(\frac{-\log[1 - (1 - e^{-\lambda})T_6^{\frac{1}{c}}]}{\lambda})) (1 - (1 - e^{-\lambda})T_6^{\frac{1}{c}})]^{\beta-1}}{(1 - e^{-\lambda})^{\beta-1} (\beta - 1) [B(\frac{1}{c}, r + 1)]^\beta}, \tag{2.6}$$

where T_6 has the beta distribution with parameters $\frac{1}{c}$ and $r\beta + 1$.

- Let $w_7(x) = [(F(x))^s(1 - (F(x))^c)^r]$ as a weight function; then from (1.1) we have

$$\begin{aligned} f_{\lambda,\theta}^{w_7}(x) &= \frac{[(F(x))^s(1 - (F(x))^c)^r] f(x)}{E_{f(x)}[(F(X))^s(1 - (F(X))^c)^r]} \\ &= \frac{c[1 - (\frac{1 - \exp(-\lambda G(x))}{1 - e^{-\lambda}})^c]^r [1 - \exp(-\lambda G(x))]^s \lambda g(x) \exp(-\lambda G(x))}{B(\frac{s+1}{c}, r + 1)(1 - e^{-\lambda})^{s+1}}. \end{aligned} \tag{2.7}$$

Also,

$$H_\beta(f^{w_7}) = \frac{1}{\beta - 1} - \frac{B(\frac{s\beta+1}{c}, r\beta + 1) E[g(G^{-1}(\frac{-\log[1 - (1 - e^{-\lambda})T_7^{\frac{1}{c}}]}{\lambda})) (1 - (1 - e^{-\lambda})T_7^{\frac{1}{c}})]^{\beta-1}}{(\beta - 1) [B(\frac{s+1}{c}, r + 1)]^\beta (\frac{c\lambda}{1 - e^{-\lambda}})^{1-\beta}}, \tag{2.8}$$

where T_7 has the beta distribution with parameters $\frac{s\beta+1}{c}$ and $r\beta + 1$. By replacing s and r with $ac - 1$ and $b - 1$, respectively, in (2.7), we have

$$f_{\lambda,\theta}^{w_8}(x) = \frac{c[1 - (\frac{1 - \exp(-\lambda G(x))}{1 - e^{-\lambda}})^c]^{b-1} [1 - \exp(-\lambda G(x))]^{ac-1} \lambda g(x) \exp(-\lambda G(x))}{B(a, b)(1 - e^{-\lambda})^{ac}}.$$

Hence,

$$\begin{aligned} H_\beta(f^{w_8}) &= \frac{1}{\beta - 1} - \frac{(\lambda c)^{\beta-1} B(\frac{(ac-1)\beta+1}{c}, (b - 1)\beta + 1)}{(1 - e^{-\lambda})^{\beta-1} (\beta - 1) [B(a, b)]^\beta} \times \\ &\quad E[g(G^{-1}(\frac{-\log[1 - (1 - e^{-\lambda})T_8^{\frac{1}{c}}]}{\lambda})) (1 - (1 - e^{-\lambda})T_8^{\frac{1}{c}})]^{\beta-1}, \end{aligned}$$

where T_8 has the beta distribution with parameters $\frac{(ac-1)\beta+1}{c}$ and $(b - 1)\beta + 1$

- When $w_9(x) = x$, X^w is said to be a length biased (or a size biased) random variable, and the density function in this case becomes

$$f_{\lambda,\theta}^{w_9}(x) = \frac{x f(x)}{E(X)} = \frac{\lambda x g(x) \exp(-\lambda G(x))}{(1 - e^{-\lambda}) E_U[Q_{\lambda,\theta}(U)]},$$

where $U = \frac{1 - \exp(-\lambda G(X))}{1 - e^{-\lambda}} \sim Uniform(0, 1)$, and $Q_{\lambda, \theta}(U) = G^{-1}(\frac{-\log(1 - (1 - e^{-\lambda})U)}{\lambda})$ is the quantile function of exp-G model. Also, in this case the Tsallis entropy is

$$H_{\beta}(f^{w_{\theta}}) = \frac{1}{\beta - 1} - \frac{\lambda^{\beta-1} E\{[g(Q_{\lambda, \theta}(U))Q_{\lambda, \theta}(U)]^{\beta-1} [1 - U(1 - e^{-\lambda})]^{\beta}\}}{(1 - e^{-\lambda})^{\beta-1} (\beta - 1) [E_U(Q_{\alpha, \tau}(U))]^{\beta}}, \quad (2.9)$$

• With $w_{10}(x) = x^r (F(x))^s (1 - F(x))^t$, we can define a class of moments, called moments probability weighted moments as

$$M_r = E[w_{10}(X)] = E\left\{ \left[G^{-1}\left(\frac{-\log(1 - (1 - e^{-\lambda})Z)}{\lambda}\right) \right]^r \right\},$$

where $Z \sim Beta(s + 1, t + 1)$. Corresponding to M_r , we have a pdf as

$$\begin{aligned} f_{\lambda, \theta}^{w_{10}}(x) &= \frac{x^r (F(x))^s (1 - F(x))^t f(x)}{M_r} \\ &= \frac{x^r [1 - \exp(-\lambda G(X))]^s [\exp(-\lambda G(X)) - e^{-\lambda}]^t \lambda g(x) \exp(-\lambda G(X))}{(1 - e^{-\lambda})^{s+t+1} M_r}. \end{aligned}$$

Also,

$$H_{\beta}(f^{w_{10}}) = \frac{1}{\beta - 1} - \frac{\lambda^{\beta-1} A}{(1 - e^{-\lambda})^{\beta-1} (\beta - 1) M_r^{\beta}}, \quad (2.10)$$

where T_{10} has the beta distribution with parameters $\beta s + 1$ and $\beta t + 1$, and

$$A = E\left\{ \left[g\left(G^{-1}\left(\frac{-\log(1 - (1 - e^{-\lambda})T_{10})}{\lambda}\right)\right) \right]^{\beta-1} \left[G^{-1}\left(\frac{-\log(1 - (1 - e^{-\lambda})T_{10})}{\lambda}\right) \right]^{\beta-1} [1 - T_{10}(1 - e^{-\lambda})]^{r\beta} \right\}.$$

3 Residual entropy for exp-G distribution

Ebrahimi (1996) considered the dynamic measure of uncertainty of the residual life random variable $(X - t | X > t)$. It is given by

$$\begin{aligned} H(f; t) &= - \int_t^{+\infty} \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)} dx \\ &= \log \bar{F}(t) - \frac{1}{\bar{F}(t)} \int_t^{+\infty} f(x) \log f(x) dx \\ &= 1 - \frac{1}{\bar{F}(t)} \int_t^{+\infty} f(x) \log \lambda_F(x) dx, \end{aligned} \quad (3.1)$$

where $\lambda_F(x) = \frac{f(x)}{\bar{F}(x)}$ is the hazard or failure function. Nanda and Paul (2006) defined residual entropy measures of order β (for $\beta > 0$, $\beta \neq 1$) as

$$H_{\beta}(f; t) = \frac{1}{\beta - 1} \left(1 - \int_t^{+\infty} \left[\frac{f(x)}{\bar{F}(t)} \right]^{\beta} dx \right), \quad H_{\beta}^*(f; t) = \frac{1}{1 - \beta} \log \left(\int_t^{+\infty} \left[\frac{f(x)}{\bar{F}(t)} \right]^{\beta} dx \right). \quad (3.2)$$

By choosing appropriate β , $H_\beta(f; t)$ and $H_\beta^*(f; t)$ can always be made non-negative. Also, It can be noted that as $\beta \rightarrow 1$, then these entropy measures reduce to residual entropy in (3.1).

Residual entropy measures of order β for exp-G model are given by

$$H_\beta(f; t) = \frac{1}{\beta - 1} \left[1 - \frac{\lambda^{\beta-1}(1 - e^{-\lambda})}{[e^{-\lambda G(t)} - e^{-\lambda}]^\beta} \int_{\frac{1-e^{-\lambda H(t)}}{1-e^{-\lambda}}}^1 \{g(Q_{\lambda, \theta}(U))[1 - U(1 - e^{-\lambda})]\}^{\beta-1} du \right], \quad (3.3)$$

and

$$H_\beta^*(f; t) = - \log \lambda - \frac{\log[(e^{-\lambda G(t)} - e^{-\lambda})^\beta (1 - t)]}{1 - \beta} + \frac{\log\left[\int_{\frac{1-e^{-\lambda H(t)}}{1-e^{-\lambda}}}^1 \{g(Q_{\lambda, \theta}(U))[1 - U(1 - e^{-\lambda})]\}^{\beta-1} du\right]}{1 - \beta}. \quad (3.4)$$

Note that when $\beta \rightarrow 1$, then (3.3) and (3.4) reduce to residual entropy of exp-G model as

$$H(f; t) = 1 - \log \lambda - \frac{(1 - e^{-\lambda})E\{\log g(X) - \lambda G(X) - \log(e^{-\lambda G(X)} - e^{-\lambda}) \mid X > t\}}{1 - e^{-\lambda G(t)}}.$$

References

- Barreto-Souza, W. and Simas, A. B. (2010). The exp-G family of probability distributions, *Brazilian Journal of Probability and Statistics*, **27(1)**, 84–109.
- Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*, John Wiley, New York.
- Ebrahimi, N. (1996). How to measure uncertainty in the residual lifetime distributions, *Sankhya*, **58**, 48–57.
- Havrda, J. and Charvat, F. (1967). Quantification method in classification processes: concept of structural K-entropy, *Kybernetika*, **3**, 30–35.
- Kapur, J.N. (1994). *Measure of Information and Their Applications*, John Wiley, New York.
- Nadarajah, S. and Zografos, K. (2003). Formulas for Renyi information and related measures for univariate distributions, *Information Sciences*, **155**, 119–138.
- Nanda, K.A. and Paul, P. (2006). Some results on generalized residual entropy, *Information Sciences*, **176**, 27–47.
- Patil, G. P. (2002). *Weighted Distributions*. In *Encyclopedia of Environmetrics*, Eds. A.H. EI-Shaarawi and W.W. Piegorsch, volume 4, pages 2369–2377, John Wiley and Sons, Ltd. Chichester.

- Rao, C. R.(1965). On discrete distributions arising out of methods of ascertainment, *Sankhy: The Indian Journal of Statistics, Series A*, **27(2/4)**, 311–324.
- Renyi, A. (1961). On measures of entropy and information, *Proc. 4th Berkeley Symposium, Stat. Probability*, **1**, 547–561.
- Riabi, M. Y. A., Mohtashami Borzadaran, G. R. and Yari, G. H. (2010). β -entropy for Pareto-type distributions and related weighted distributions, *Statistics and Probability Letters*, **80**, 1512–1519.
- Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379–432.
- Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics*, **52(1/2)**, 479–487.



Extreme Value Distributions for Skew-t Distributions

Asma Taimouri¹, Mohsen Khosravi¹, Ahad Jamalizadeh¹

¹Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman,
Kerman, Iran

Abstract: In this paper we discuss extreme value distributions of several skew-t distributions. Recently, Chang and Genton (2007) obtained extreme value distributions for the three types of skew-t distributions. Here we modified some of their results about Type III skew-t distribution. We consider an extension of Type I skew-t distribution, termed as skew-t-t distribution, and we obtain its extreme value distribution. Finally, we derive the extreme value distribution of the skew-t-normal distribution.

Keywords Domain of attraction, Extreme value, Skew-t distribution, Skew-t-normal distribution, Tail ratio.

Mathematics Subject Classification (2010): 60G70 62H10 62E10.

1 Introduction

1.1 Skew-t Distributions

The skew-t distribution is an asymmetric probability model that is receiving considerable attention, see for example [Azzalini and Capitanio \(2003\)](#) and [Jones and Faddy \(2003\)](#). Let $s(x; \nu)$ and $S(x; \nu)$ denote the pdf and cdf, respectively, of the t-distribution with ν degrees of freedom. There are three popular definitions of skew-t distributions. The Type I skew-t ($ST - I$) implicit from [Azzalini \(1985\)](#) with pdf

$$g_I(x; \nu, \lambda) = 2s(x; \nu)S(\lambda x; \nu), \quad (1.1)$$

where $\lambda \in \mathbb{R}$. When $\nu = 1$, g_I is the pdf of the skew-Cauchy distribution ([Arnold and Beaver \(2000\)](#)). The Type II skew-t ($ST - II$) proposed by [Branco and Dey \(2001\)](#), and in an equivalent form by [Azzalini and Capitanio \(2003\)](#), with pdf

$$g_{II}(x; \nu, \lambda) = 2s(x; \nu)S\left(\lambda x \sqrt{\frac{1 + \nu}{x^2 + \nu}}; \nu + 1\right), \quad (1.2)$$

¹Asma Taimouri : Taimouri.asma@yahoo.com

and the Type III skew-t ($ST - III$) proposed by Jones (2001) with pdf

$$g_{III}(x; \eta_a, \eta_b) = \Delta_{\eta_a, \eta_b}^{-1} \left(1 + \frac{x}{[\eta_a + \eta_b + x^2]^{\frac{1}{2}}}\right)^{\eta_a + \frac{1}{2}} \left(1 - \frac{x}{[\eta_a + \eta_b + x^2]^{\frac{1}{2}}}\right)^{\eta_b + \frac{1}{2}} \quad (1.3)$$

where $\eta_a, \eta_b > 0$, $\Delta_{\eta_a, \eta_b} = 2^{\eta_a + \eta_b - 1} (\eta_a + \eta_b)^{1/2} B(\eta_a, \eta_b)$ and B is the beta function.

1.2 Extreme Value Theory

Extreme value theory and its relation to order statistics have different applications in various fields, see for example Embrechts et al. (1997), Kotz and Nadarajah (2000) and Coles (2001).

Let X_1, \dots, X_n be independent and identically distributed (i.i.d) random variables with cdf F and let $M_n = \max_{i=1, \dots, n} X_i$ denote the sample maximum. We focus on the maximum since the minimum is defined by $\min_{i=1, \dots, n} X_i = -\max_{i=1, \dots, n} (-X_i)$, so the sample minimum from the cdf $F_{(1)}$ has the sample distribution as the negative of the sample maximum from the cdf $F_{(n)}$ where $F_{(n)}(x) = 1 - F_{(1)}(-x)$, then the results can be reformulated for the sample minimum.

If sequences of scale and location normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ exist such that the limiting distribution

$$\lim_{n \rightarrow \infty} Pr \left(\frac{M_n - b_n}{a_n} \leq x \right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \quad (1.4)$$

is non-degenerate, then the limit G is an *Extreme - Value (EV)* distribution.

The Fisher-Tippett theorem plays a key role in extreme value theory and is as follows:

Theorem 1.1. *If (1.4) holds, the limiting cdf G is one of the following three families of extreme value distributions (EVD's):*

- (1) *Gumbel* : $\Lambda(x) = \exp(-e^{-x})$, $-\infty < x < +\infty$,
- (2) *Frechet* : $\Phi_a(x) = \exp(-x^{-a})$, $x > 0$, $a > 0$,
- (3) *Weibull* : $\Psi_a(x) = \exp(-(-x)^a)$, $x \leq 0$, $a > 0$.

Sketches of proofs, extensions, choices of normalizing constants, and applications, can be found in Embrechts et al. (1997), Kotz and Nadarajah (2000) and Coles (2001).

In applications, the problem is to identify the family to which G belongs. Embrechts et al. (1997), Resnick (1987) and Leadbetter et al. (1983) summarize how this is done.

Let $D(G)$ denote the domain of attraction of G , i.e., the collection of all those cdf's for which the distribution of the sample maximum converges to G (where G is one of the three EVD's: Λ , Φ_a , or Ψ_a). Following Resnick (1987), define the tail ratio of a cdf F as

$$\lim_{n \rightarrow \infty} \frac{1 - F(tx)}{1 - F(x)}, \quad (1.5)$$

where $t > 1$. Then $F \in D(\Phi_a)$ if and only if (1.5) converges to t^{-a} with $a > 0$. If $F \in D(\Lambda)$, then the limit (1.5) is equal to 0. Note that if F is continuous and defined on \mathbb{R} , then F never belongs to $D(\Psi_a)$.

The remainder of the paper is organized as follows. In section 2, we modify the extreme value distribution of Type III skew-t distribution. In section 3, we consider an extension of Type I skew-t distribution, and we drive new results of the extreme value distributions for generalized Type I skew-t (skew-t-t) and skew-t-normal distributions.

2 The Extreme Value Distribution of Type III Skew-t Distribution

Chang and Genton (2007) obtained the EVD’s of three type skew-t distributions. They showed that Type I skew-t distribution with cdf G_I belongs to $D(\Phi_a)$ with $a = \nu$ when $\alpha \geq 0$ and $a = 2\nu$ when $\alpha < 0$, the Type II skew-t distribution with cdf G_{II} belongs to $D(\Phi_a)$ with $a = \nu$ for all $\alpha \in \mathbb{R}$ and the Type III skew-t distribution with cdf G_{III} belongs to $D(\Phi_a)$ with $a = 2\eta_a$ for all $\eta_a, \eta_b > 0$, in which that the result of Type III skew-t distribution must be modified.

Theorem 2.1. *The Type III skew-t distribution with pdf’s g_{III} and associated cdf’s G_{III} have the following EVD’s:*

$$G_{III}(x; \eta_a, \eta_b) \text{ is in } D(\Phi_a) \text{ with } a = 2\eta_b \text{ for all } \eta_a, \eta_b > 0.$$

Proof. In order to relate the $ST - III$ to results already obtained, its pdf can be rewritten as

$$\begin{aligned} g_{III}(x; \eta_a, \eta_b) &= \Delta_{\eta_a, \eta_b}^{-1} \left(1 + \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{\eta_a + \frac{1}{2}} \left(1 - \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{\eta_b + \frac{1}{2}} \\ &= \Delta_{\eta_a, \eta_b}^{-1} \left(1 + \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{\eta_a + \frac{1}{2}} \left(1 - \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{\eta_b + \frac{1}{2}} \\ &\quad \times \left(1 + \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{\eta_b + \frac{1}{2}} \left(1 + \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{-(\eta_b + \frac{1}{2})} \\ &= \Delta_{\eta_a, \eta_b}^{-1} \left(1 + \frac{x^2}{\eta_a + \eta_b} \right)^{-\frac{\eta_a + \eta_b + 1}{2}} \left(1 + \frac{x^2}{\eta_a + \eta_b} \right)^{-\frac{\eta_b - \eta_a}{2}} \\ &\quad \times \left(1 + \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{\eta_a + \frac{1}{2}} \left(1 + \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{-(\eta_b + \frac{1}{2})} \end{aligned}$$

$$= \Delta_{\eta_a, \eta_b}^{-1} \left(1 + \frac{x^2}{\eta_a + \eta_b} \right)^{-\frac{\eta_a + \eta_b + 1}{2}} \times \Pi^*(x, \eta_a, \eta_b),$$

where

$$\Pi^*(x, \eta_a, \eta_b) = \left(1 + \frac{x^2}{\eta_a + \eta_b} \right)^{-\frac{\eta_b - \eta_a}{2}} \left(1 + \frac{x}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}}} \right)^{\eta_a - \eta_b} = \left(\frac{(\eta_a + \eta_b)^{\frac{1}{2}}}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}} + x} \right)^{\eta_b - \eta_a},$$

and we have

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{\Pi^*(tx, \eta_a, \eta_b)}{\Pi^*(x, \eta_a, \eta_b)} &= \lim_{x \rightarrow \infty} \frac{\left(\frac{(\eta_a + \eta_b)^{\frac{1}{2}}}{(\eta_a + \eta_b + t^2 x^2)^{\frac{1}{2}} + tx} \right)^{\eta_b - \eta_a}}{\left(\frac{(\eta_a + \eta_b)^{\frac{1}{2}}}{(\eta_a + \eta_b + x^2)^{\frac{1}{2}} + x} \right)^{\eta_b - \eta_a}} \\ &= \lim_{x \rightarrow \infty} \left(\frac{(x^2)^{\frac{1}{2}} + x}{(t^2 x^2)^{\frac{1}{2}} + tx} \right)^{\eta_b - \eta_a} = \lim_{x \rightarrow \infty} \left(\frac{1}{t} \right)^{\eta_b - \eta_a} = t^{\eta_a - \eta_b}. \end{aligned}$$

Then, for any $t > 1$, the tail ratio of G_{III} is

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{1 - G_{III}(tx; \eta_a, \eta_b)}{1 - G_{III}(x; \eta_a, \eta_b)} &= \lim_{x \rightarrow \infty} \frac{tg_{III}(tx; \eta_a, \eta_b)}{g_{III}(x; \eta_a, \eta_b)} \\ &= \lim_{x \rightarrow \infty} \frac{t \left(1 + \frac{t^2 x^2}{\eta_a + \eta_b} \right)^{-\frac{\eta_a + \eta_b + 1}{2}}}{\left(1 + \frac{x^2}{\eta_a + \eta_b} \right)^{-\frac{\eta_a + \eta_b + 1}{2}}} \times \frac{\Pi^*(tx, \eta_a, \eta_b)}{\Pi^*(x, \eta_a, \eta_b)} \\ &= \lim_{x \rightarrow \infty} t^{-(\eta_a + \eta_b)} \times \lim_{x \rightarrow \infty} \frac{\Pi^*(tx, \eta_a, \eta_b)}{\Pi^*(x, \eta_a, \eta_b)} \\ &= t^{-(\eta_a + \eta_b)} \times t^{\eta_a - \eta_b} = t^{-2\eta_b}. \end{aligned}$$

That is, G_{III} belongs to $D(\Phi_a)$ with $a = 2\eta_b$ and the proof is complete. \square

3 Extreme value Distributions for Generalized Type I Skew-t and Skew-t-Normal Distributions

In this section, we discuss the EVD's of generalized Type I skew-t and skew-t-normal distributions. We use theorems 3.1-3.3 of Chang and Genton (2007) to determine whether G belongs to $D(\Lambda)$ or $D(\Phi_a)$.

Wang et al. (2004) defined skew-symmetric (*SS*) distributions by means of their probability density function (pdf), *g* say, given by

$$g(x) = 2f(x)\pi(x), \tag{3.1}$$

where $f : \mathbb{R} \rightarrow [0, +\infty)$ is a symmetric pdf, and $\pi : \mathbb{R} \rightarrow [0, 1]$ is a skewing function satisfying $\pi(x) + \pi(-x) = 1$. When $f(x) = s(x; \nu)$, the standard t-distribution pdf and $\pi = \Phi$, the standard normal cdf, the distribution with pdf (3.1) is referred to as being skew-t-normal distribution. Assume that;

- (A) π is continuous and there exists a constant $M > 0$ such that $\pi(x)$ is monotone for $x > M$,
- (B) f and π have continuous second derivatives,
- (C) there exists a constant $M^* > 0$ such that $g'(x) < 0$ for $x > M^*$.

Theorem 3.1. (*Chang and Genton (2007)*) Let f be a symmetric pdf with associated cdf F , and g the corresponding skew-symmetric pdf defined by (3.1) with associated cdf G and skewing function π . Assume (A), (B) and (C). If $F \in D(Q)$ and

$$\lim_{x \rightarrow \infty} \pi(x) = \omega \in (0, 1],$$

then $G \in D(Q)$ also, where Q is either Λ (Gumbel) or Φ_a (Frechet).

Theorem 3.2. (*Chang and Genton (2007)*) Assume (A), (B) and (C). Then, $G \in D(\Lambda)$ if

$$\lim_{x \rightarrow \infty} \frac{[1 - G(x)]G''(x)}{[G'(x)]^2} = -1. \tag{3.2}$$

Theorem 3.3. (*Chang and Genton (2007)*) Assume (A), (B) and (C). Suppose $F \in D(\Phi_a)$ and $\lim_{x \rightarrow \infty} \pi(x) = 0$. If, for any $t > 1$,

$$\lim_{x \rightarrow \infty} \frac{\pi(tx)}{\pi(x)} > 0, \tag{3.3}$$

then $G \in D(\Phi_a)$.

Here, we consider a generalized Type I skew-t (skew-t-t) distribution with the following pdf

$$g_{STT}(x; \nu_1, \nu_2) = 2s(x; \nu_1)S(\lambda x; \nu_2) \quad x \in \mathbb{R}, \quad \lambda \in \mathbb{R}, \quad \nu_1, \nu_2 > 0. \tag{3.4}$$

Chang and Genton (2007) obtained the EVD of the Type I skew-t distribution with pdf (1.1). Here, by substituting ν_1 and ν_2 instead of ν , in the pdf (1.1), we drive the EVD of the skew-t-t distribution.

Theorem 3.4. The skew-t-t distribution with pdf (3.4) and associated cdf G_{STT} is belong to $D(\Phi_a)$ with $a = \nu_1$ when $\lambda \geq 0$ and $a = \nu_1 + \nu_2$ when $\lambda < 0$.

Proof. If $\lambda \geq 0$, we have $\lim_{x \rightarrow \infty} \pi(x) = \lim_{x \rightarrow \infty} S(\lambda x; \nu) = 1 \in (0, 1]$.

On the other side, the tail ratio of $S(x; \nu_1)$ is $t^{-\nu_1}$, i.e., $S(x; \nu_1) \in D(\Phi_a)$ with $a = \nu_1$. According to theorem 3.1 we have, $G_{STT} \in D(\Phi_a)$ with $a = \nu_1$ when $\lambda \geq 0$.

When $\lambda < 0$, we have $\lim_{x \rightarrow \infty} \pi(x) = \lim_{x \rightarrow \infty} S(\lambda x; \nu) = 0$.

On the other hand, we have, by L'Hospital's rule,

$$\lim_{x \rightarrow \infty} \frac{S(\lambda tx; \nu_2)}{S(\lambda x; \nu_2)} = \lim_{x \rightarrow \infty} \frac{ts(\lambda tx; \nu_2)}{s(\lambda x; \nu_2)} = \lim_{x \rightarrow \infty} \frac{t(\nu_2 + t^2 \lambda^2 x^2)^{-\frac{\nu_2+1}{2}}}{(\nu_2 + \lambda^2 x^2)^{-\frac{\nu_2+1}{2}}} = \lim_{x \rightarrow \infty} t^{-\nu_2} = t^{-\nu_2},$$

and also we have

$$\lim_{x \rightarrow \infty} \frac{s(\lambda tx; \nu_1)}{s(\lambda x; \nu_1)} = \lim_{x \rightarrow \infty} t^{-\nu_1} = t^{-\nu_1}.$$

Therefore, the tail ratio of W is

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{1 - G_{STT}(tx; \nu_1, \nu_2)}{1 - G_{STT}(x; \nu_1, \nu_2)} &= \lim_{x \rightarrow \infty} \frac{g_{STT}(tx; \nu_1, \nu_2)}{g_{STT}(x; \nu_1, \nu_2)} = \lim_{x \rightarrow \infty} \frac{s(tx; \nu_1)}{s(x; \nu_1)} \times \lim_{x \rightarrow \infty} \frac{S(\lambda tx; \nu_2)}{S(\lambda x; \nu_2)} \\ &= t^{-\nu_1} \times t^{-\nu_2} = t^{-(\nu_1 + \nu_2)} = t^{-a}; \quad a = \nu_1 + \nu_2. \end{aligned}$$

That is, by theorem 3.3 $G_{STT} \in D(\Phi_a)$ with $a = \nu_1 + \nu_2$ for $\lambda < 0$. □

Gomez et al. (2007) discussed a skew-t-normal (STN) distribution with the following pdf

$$g_{STN}(x; \nu, \lambda) = 2s(x; \nu)\Phi(\lambda x), \quad x \in \mathbb{R}, \quad \lambda \in \mathbb{R}, \quad \nu > 0. \tag{3.5}$$

Ho et al. (2013) and Lin et al. (2014) used this distribution for mixture modeling. It is well known that $\Phi \in D(\Lambda)$. The tail ratio of S is $t^{-\nu}$, i.e., $S \in D(\Phi_a)$ with $a = \nu$. In the following theorem we derive the EVD of the STN distribution.

Theorem 3.5. *Let (3.5) be the pdf of a STN distribution with associated cdf G_{STN} . Then, G_{STN} is in $D(\Phi_a)$ with $a = \nu$ when $\lambda \geq 0$ and G_{STN} is in $D(\Lambda)$ when $\lambda < 0$.*

Proof. Considering the tail ratio of the STN distribution with cdf G_{STN} , we have

$$\lim_{x \rightarrow \infty} \frac{1 - G_{STN}(tx; \nu)}{1 - G_{STN}(x; \nu)} = \lim_{x \rightarrow \infty} \frac{tg_{STN}(tx; \nu)}{g_{STN}(x; \nu)} = \lim_{x \rightarrow \infty} \frac{ts(tx; \nu)}{s(x; \nu)} \times \lim_{x \rightarrow \infty} \frac{\Phi(\lambda tx)}{\Phi(\lambda x)} \tag{3.6}$$

for $t > 1$. On the other hand, if $\lambda \geq 0$ we have $\lim_{x \rightarrow \infty} \frac{\Phi(\lambda tx)}{\Phi(\lambda x)} = 1$, and if $\lambda < 0$ since $t > 1$ and $t^2 - 1 > 0$ we have

$$\lim_{x \rightarrow \infty} \frac{\Phi(\alpha tx)}{\Phi(\alpha x)} = \lim_{x \rightarrow \infty} \frac{\alpha t \phi(\alpha tx)}{\alpha \phi(\alpha x)} = \lim_{x \rightarrow \infty} te^{-\frac{\alpha^2 x^2}{2}(t^2 - 1)} = 0.$$

Since the tail ratio of the t-distribution is $t^{-\nu}$, (3.6) is equal to $t^{-\nu}$ if λ is non negative; thus $G_{STN} \in D(\Phi_a)$ with $a = \nu$ when $\lambda \geq 0$. But since (3.6) for $\lambda < 0$ is equal to zero, G_{STN} with negative λ is not in $D(\Phi_a)$. According to theorem 3.2 we can show that G_{STN} is in $D(\Lambda)$ for $\lambda < 0$. Now, when x is large and $\lambda < 0$, we have $1 - \Phi(x) \approx \frac{\phi(x)}{x}$, therefore we have $G'_{STN}(x; \nu, \lambda) = g_{STN}(x; \nu, \lambda) \approx -2s(x; \nu) \frac{\phi(\lambda x)}{\lambda x}$, $G''_{STN}(x; \nu, \lambda) = g'_{STN}(x; \nu, \lambda) \approx -\frac{2}{\lambda x} [s'(x; \nu) - \lambda^2 x s(x; \nu)] \phi(\lambda x)$ and $G'''_{STN}(x; \nu, \lambda) = g''_{STN}(x; \nu, \lambda) \approx -\frac{2}{\lambda x} \phi(\lambda x) [s''(x; \nu) - 2\lambda^2 x s'(x; \nu) + \lambda^4 x^2 s(x; \nu)]$.

Also $\lim_{x \rightarrow \infty} \frac{G''_{STN}(x; \nu, \lambda)}{x G'_{STN}(x; \nu, \lambda)} = -\lambda^2$ and $\lim_{x \rightarrow \infty} \frac{1 - G_{STN}(x; \nu, \lambda)}{G''_{STN}(x; \nu, \lambda)} = \lim_{x \rightarrow \infty} \frac{s(x; \nu)}{\lambda^2 x s(x; \nu)} = 0$. Thus,

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{[1 - G_{STN}(x; \nu, \lambda)] G''_{STN}(x; \nu, \lambda)}{[G'_{STN}(x; \nu, \lambda)]^2} &= \lim_{x \rightarrow \infty} \frac{[1 - G_{STN}(x; \nu, \lambda)] x}{G'_{STN}(x; \nu, \lambda)} \times \frac{G''_{STN}(x; \nu, \lambda)}{x G'_{STN}(x; \nu, \lambda)} \\ &= \lim_{x \rightarrow \infty} \frac{[1 - G_{STN}(x; \nu, \lambda)] x}{G'_{STN}(x; \nu, \lambda)} \times \lim_{x \rightarrow \infty} \frac{G''_{STN}(x; \nu, \lambda)}{x G'_{STN}(x; \nu, \lambda)} \\ &= [0 - (-\frac{1}{\lambda^2})] \times (-\lambda^2) = -1. \end{aligned}$$

So the condition of theorem 3.2 is satisfied and hence G_{STN} is in $D(\Lambda)$. □

References

Arnold, B. C., Beaver, R. J. (2000). The skew-Cauchy distribution. *Statist. Probab. Lett.* 49:285290.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* 12:171178.

Azzalini, A., Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *J. Roy. Stat. Soc. Ser. B.* 65:367389.

Branco, M. D., Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.* 79:99113.

Chang, S., Genton, M. G. (2007). Extreme value distributions for the skew-symmetric family of distributions. *Communications in Statistics Theory and Methods.* 36:17051717.

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. *New York: Springer.*

Embrechts, P., Kluppelberg, C., Mikosch, T. (1997). Modelling Extremal Events: For Insurance and Finance. *New York: Springer.*

- Fisher, R. A., Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philosophical Soc.* 24:180-190.
- Gomez, H. W., Venegas, O., Bolfarin, H., (2007). Skew-Symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics.* 18:395-407.
- Ho, H. J., pyne, S., Lin, T. I. (2012b). Maximum likelihood inference for mixtures of skew student-t-normal distributions through practical EM-type algorithms. *stat. comput.* 22:287-299.
- Jones, M. C. (2001). Multivariate t and beta distributions associated with the multivariate F distribution. *Metrika.* 54:215-231.
- Jones, M. C., Faddy, M. J. (2003). A skew extension of the t distribution, with applications. *J. Roy. Stat. Soc. Ser. B* 65:159-174.
- Kotz, S., Nadarajah, S. (2000). Extreme Value Distributions: Theory and Applications. *U.K: Imperial College Press.*
- Leadbetter, M. R., Lindgren, G., Rootzn, H. (1983). Extremes and Related Properties of Random Sequences and Processes. *New York: Springer.*
- Lin, T. I. (2010). Robust mixture modeling using multivariate skew-t distribution. *state. compute.* 20:343-356.
- Resnick, S. I. (1987). Extreme Values, Regular Variation, and Point Processes. *New York: Springer-Verlag.*



Hypothesis Testing of Hidden Periodicities by Quantile Periodogram

M. TAZIKINEJAD¹, M. KHALAFI², M. AZIMMOHSENI³
Golestan University

Abstract: Detecting hidden periodicities in a time series is a challenging problem. Fisher's test based on the periodogram, a common method for this purpose, is not a powerful tool to detect all effective hidden periodicities. The quantile periodogram introduced by Li (2012) is an alternative approach for this purpose. In this article, based on a simulation study, we try to show the performance of the quantile periodogram.

Keywords Hidden periodicity; Quantile periodogram; Quantile regression

Mathematics Subject Classification (2010): 62XX 62MXX 62M15.

1 Introduction

From the fact that a periodic time series can be assumed as a regression model with trigonometric explanatory variables, the quantile periodogram is introduced as a special case of the quantile regression. The quantile regression, introduced by Koenker and Bassett (1978), shows a powerful technique for analyzing the quantiles of data that are important for many applied studies, (Koenker 2005). This technique may be viewed as an extension of classical least squares estimation of conditional mean models to the estimation of an ensemble of models for several conditional quantile functions.

To see the quantile regression from the theoretical point of view one may refer to (Ying et al 2009), and to find its applications one may refer to (Ming et al 2013).

The analytical results demonstrate the capacity of the quantile periodogram for detecting hidden periodicities in the quantiles and characterizing time series with time dependent variance (Li, 2014). For this reason, recent studies on this topic are mainly regarded to other capacities of this method. For instance one may refer to the work of (Kim, 2015) finding quantile regression as a powerful method to detect of ARCH models by combining distributional information across multiple quantile regression. Li (2012) shows the performance of quantile periodogram in compare to Fisher's test for testing the null hypothesis of a time series is Gaussian white noise versus the hypothesis that the data generated by a Gaussian white noise sequence with a superimposed deterministic periodic component. Indeed,

³Marziye Tazikinejad : marziyetaziki@yahoo.com

he shows that Fisher's test unable to detect hidden periodicity in the standard deviation.

Although the performance of the quantile periodogram is proved from a theoretical point of view, but in the present manuscript we try to find more features of this method by conducting a numerical study based on simulated data for different cases. In each case, we compare quantile and ordinary periodograms to detect the hidden periodicities.

This manuscript organized as follows. First, we review the concept of quantile periodogram. Afterward, in section 3, we perform an extensive simulation study to compute the power of quantile periodogram to detect hidden predetermined hidden periodicities.

2 Quantile periodogram

An application of quantile regression in the field of time series analysis introduced by Li (2012), to take a trigonometric regressor $x_t(\omega) := (\cos(\omega t), \sin(\omega t))^T$ and consider the linear trigonometric quantile regression solution,

$$\{\hat{\beta}_{n,\alpha}(\omega)\} := \arg \min_{\beta \in \mathbb{R}e^2} \sum_{t=1}^n \rho_{\alpha}(Y_t - \lambda - x_t^T(\omega)\beta), \quad (1)$$

where $\omega := 2\pi f \in (0, \pi)$ is the frequency variable and λ is a suitable constant, typically the α -quantile of $\{Y_t\}$ and the loss function $\rho_{\alpha}(u)$ defined as

$$\rho_{\alpha}(u) := \begin{cases} -(1 - \alpha)u & u < 0 \\ \alpha u & u \geq 0 \end{cases} .$$

So, the quantile periodogram is computed as,

$$Q_{n,\alpha}(\omega) := \frac{1}{4} n \|\hat{\beta}_{n,\alpha}(\omega)\|^2 \quad (2)$$

Since the statistical properties of Quantile periodogram require the asymptotic distribution of the quantile periodogram, then expressed its distribution.

Asymptotic Theory:

Assuming that $\{Y_t\}$ is a time series, let $F_t(u)$ and $F_{ts}(u, v)$ denote the univariate and bivariate marginal cumulative distribution function of $\{Y_t\}$,

respectively, and let the following conditions be satisfied for fixed α (and hence λ)

(C1) $f_t(u) := \dot{F}_t(u)$ exists for all u and $F_t(u + \lambda) - F_t(\lambda) = f_t(\lambda)u + O(u^{d+1})$ uniformly for $|u| \leq u_0$.

(C2) $F_t(\lambda) = \alpha$ and $f_t(\lambda) = \kappa > 0$ for all t .

(C3) $\{Y_t\}$ is stationary in λ -level crossings, that is, $P\{(Y_t - \lambda)(Y_s - \lambda) < 0\} = \gamma_{t-s}$ for all t and s .

We call γ_τ the lag- τ level-crossing rate (which depends on λ , of course).

(C4) $\{Y_t\}$ is an m -dependence process or a linear process of the form $\lambda + \sum_{i=-\infty}^{\infty} \varphi_i e_{t-i}$, where $\{e_t\}$ is an iid random sequence with $E(|e_t|) < \infty$ and $\{\varphi_t\}$ is an absolutely summable deterministic sequence such that

$$\sum_{|\iota| > n^r} \varphi_t = o(n^{-1}) \text{ as } n \rightarrow \infty \text{ for some constant } r \in [0, 1/4].$$

The sequence $\{\gamma_\tau\}$ can be represented in the frequency domain by the Fourier transform

$$\begin{aligned} S(\omega) &:= \sum_{\tau=-\infty}^{\infty} \left\{1 - \frac{1}{2\alpha(1-\alpha)} \gamma_\tau\right\} \exp(i\omega\tau) \\ &= \sum_{\tau=-\infty}^{\infty} \left\{1 - \frac{1}{2\alpha(1-\alpha)} \gamma_\tau\right\} \cos(\omega\tau) \end{aligned}$$

we call $S(\omega)$ the level-crossing spectrum of $\{Y_t\}$.

Theorem 2.1. *Let $\{Y_t\}$ satisfy the conditions (C1)-(C4). Let $Q_{n,\alpha}(\omega_j)$ ($j = 1, \dots, q$) be defined by (2) with $\hat{\beta}_{n,\alpha}(\omega)$ given by (1). Assume further that $S(\omega)$ is finite and bounded away from zero. then, as $n \rightarrow \infty$,*

$$\{Q_{n,\alpha}(\omega_j)\} \stackrel{A}{\sim} \{(1/2)\eta_1^2 S(\omega_j)\xi_j\}$$

where the symbol $\stackrel{A}{\sim}$ stands for "asymptotically distributed as", the sequence $\{\xi_j\}$ contain iid central chi-square random variables with two degree of freedom, denoted as χ_2^2 , and where $\eta_1^2 := \alpha(1-\alpha)/\kappa^2$ is scaling constant.

3 Application

In this section we provide some simulation examples to demonstrate the benefit of the quantile periodogram over the ordinary periodogram.

3.1 Detecting hidden periodicity at the time series

Assume $Y_t = a_t X_t \quad (t = 1, \dots, n) \tag{3}$

where $\{X_t\}$ is a stationary random process and $\{a_t\}$ is a nonnegative periodic sequence of deterministic of the form

$$a_t = b_0 + b_1 \cos(\omega_0 t) + b_2 \sin(\omega_0 t) \quad (0 < \omega_0 < \pi)$$

for a time series which is simulated according to (3) with $n=300$ ($b_0 = 1, b_1 = 0.9, b_2 = 0.8$, and $\omega_0 = 2\pi \times 0.1$) and with $\{X_t\}$ being a second-order autoregressive process satisfying

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t \quad (t = 1, \dots, n)$$

where $a_1 = 0$ and $a_2 = -0.36$ and $\{\varepsilon_t\}$ is Gaussian white noise.

As we can see from Figure 1, the quantile periodogram successfully reveal the hidden periodicity as a large spike at frequency ω_0 , whereas the ordinary periodogram fail to do so.

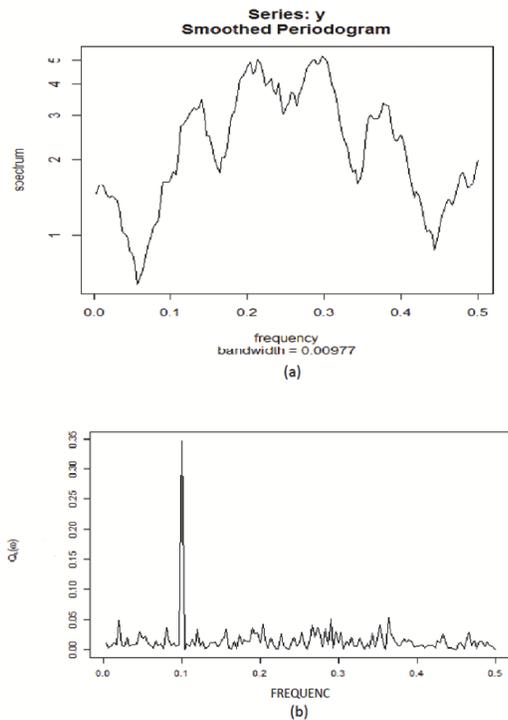


Figure 1: (a) Smooth Periodogram, (b) Quantile Periodogram with $\alpha = 0.9$

3.2 Characterizing time series with time-dependent variance

Assume $Y_t = \sigma_t Z_t$ ($t = 1, \dots, n$)

$\sigma_t = 1 + 1/\{1 + 0.00002(t - 50)^4\} + 1/\{1 + 0.00002(t - 160)^4\}$ and $n=200$ and $\{Z_t\}$ is assumed to be Gaussian white noise.

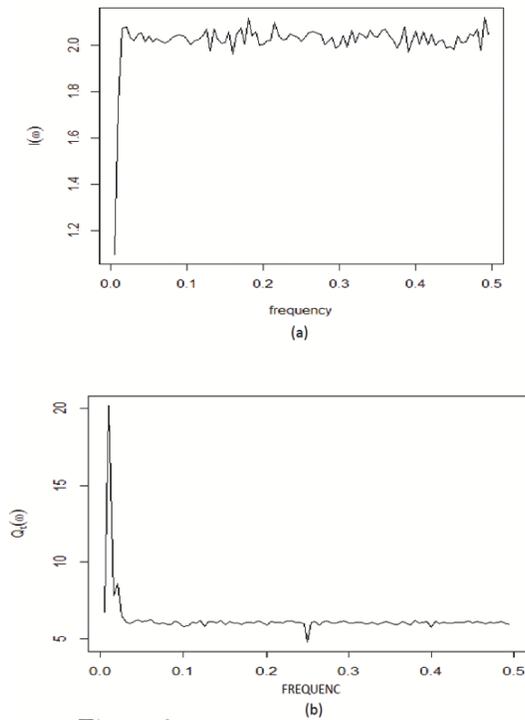


Figure 2: (a) Ordinary Periodogram, (b) Quantile Periodogram with $\alpha = 0.9$

As we can see from Figure 2 (a) the ordinary periodogram is unable to characterize the time-dependent variance for the time series while in Figure 2 (b) the quantile periodogram well depicts non-stationary time series.

3.3 Fisher's Test for Hidden Periodicities

By replacing the ordinary periodogram with the quantile periodogram, the resulting statistic takes the form

$$g_{n,\alpha} := \frac{\max\{Q_{n,\alpha}(\omega_l)\}}{\sum Q_{n,\alpha}(\omega_l)}$$

$\theta_{n,\alpha}$ is an empirical threshold which is obtained from a simulation under the null hypothesis of Gaussian white noise. The null hypothesis is rejected in favour of a hidden periodicity if $g_{n,\alpha} > \theta_{n,\alpha}$.

Assum $Y_t = \sigma_t Z_t$ ($t = 1, \dots, n$)
 $\sigma_t = 1 + a_0 \cos(\omega_0 t)$ with $a_0 = 0.6$ and $\omega_0 = 2\pi \times 0.1$,

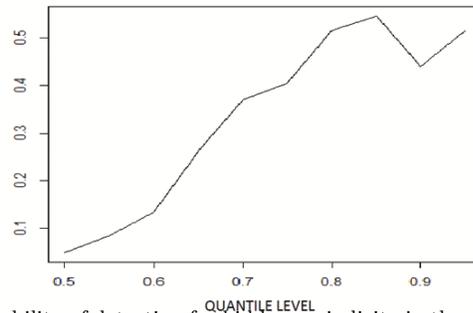


Figure 3: Probability of detection for hidden periodicity in the standard deviation by Fisher's test based on the quantile periodogram at different quantile levels.

Figure 3 depicts the probability of detection as a function of α for sample sizes 200, where the probability of false alarm is set at 0.05. As expected, the probability of detection grows at first as α increases from 0.5; the probability of detection reaches its maximum when α lies somewhere between 0.8 and 0.85; when α takes larger values, the probability of detection begins to deteriorate. A similar behaviour is observed when α moves from 0.5 toward 0 (not shown).

References

- Brockwell, P. J., and Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd ed.), New York: Springer.
- Kim, S. (2015). Hypothesis Testing for ARCH Models: A Multiple Quantile Regressions Approach, *Journal of Time Series Analysis*, **36**: 26-38.
- Koenker, K. (2005). *Quantile Regression*, Cambridge: Cambridge University Press.
- Li, TH. (2008). Laplace Periodogram for Time Series Analysis, *Journal of the American Statistical Association*, **103**: 757-768.
- Li, TH. (2012). Quantile periodograms, *Journal of the American Statistical Association*, **107**: 765-776.
- Li, TH. (2014). Quantile periodogram and time-dependent variance, *Journal of Time Series Analysis*, **35**: 322-340.
- Wei, Y. and Carroll, R. J. (2009). Quantile regression with Measurement Error, *Journal of the American Statistical Association*, **104**: 1129-1143.
- Zheng, M., Zhao, Z. and Yu, W. (2013). Quantile regression analysis of case-cohort data, *Journal of Multivariate Analysis*, **122**: 20-34.



Goodness-of-Fit Test for Normality Based on Empirical Distribution Function

Hamzeh Torabi and Narges Montazeri Hedesh²
Department of Statistics, Yazd University, Yazd, Iran.

Abstract: In this paper, a goodness-of-fit test for normality based on the comparison of the theoretical and empirical distributions is proposed. Critical values are obtained via Monte Carlo for several sample sizes and different significance levels. From the simulation study results it is concluded that the best performance against asymmetric alternatives with support on $(-\infty, \infty)$ and alternative distributions with support on $(0, \infty)$ is achieved by the new test. Other findings derived from the simulation study are that SJ and Robust Jarque-Bera tests are the most powerful ones for symmetric alternatives with support on $(-\infty, \infty)$, whereas entropy-based tests are preferable for alternatives with support on $(0, 1)$.

Keywords goodness-of-fit tests, Robust Jarque-Bera test, SJ test, test for normality.

Mathematics Subject Classification (2010): 62F03 62F10.

1 Introduction

Let X_1, \dots, X_n be a n independent and identically distributed (iid) random variables with continuous cumulative distribution function (cdf) $F(\cdot)$ and probability density function (pdf) $f(\cdot)$. All along the paper, we will denote the order statistic by $(X_{(1)}, \dots, X_{(n)})$. Based on the observed sample x_1, \dots, x_n , we are interested in the following goodness-of-fit test for a location-scale family:

$$\begin{cases} H_0 : F \in \mathcal{F} \\ H_1 : F \notin \mathcal{F} \end{cases} \quad (1.1)$$

where $\mathcal{F} = \{F_0(\cdot; \theta) = F_0\left(\frac{x-\mu}{\sigma}\right) \mid \theta = (\mu, \sigma) \in \Theta\}$, $\Theta = \mathbb{R} \times (0, \infty)$ and μ and σ are unspecified. The family \mathcal{F} is called location-scale family, where $F_0(\cdot)$ is the standard case for $F_0(\cdot; \theta)$ for $\theta = (0, 1)$. Suppose that $f_0(x; \theta) = \frac{1}{\sigma} f_0\left(\frac{x-\mu}{\sigma}\right)$ is the corresponding pdf of $F_0(x; \theta)$.

In this paper we propose a goodness-of-fit statistic to test (1.1) by considering a new distance measure between two continuous cdf's. The organization of the paper is as follows. In Section 2 we define the new distance measure H_n and study its properties as a goodness-of-fit statistic. In Section 3 we propose a normality test based on H_n and find its critical values for several sample sizes and different significance levels. In Section 4 we compare their performances to that of our proposal.

²Speaker's Name : nmontazeri@stu.yazd.ac.ir

2 A new distance measure

In this section we define a distance measure between two continuous cdf's and study its properties as a goodness-of-fit statistic.

Definition 2.1. *Let X and Y be two absolutely continuous random variables with cdf's F_0 and F , respectively. We define*

$$D(F_0, F) = \int_{-\infty}^{\infty} h\left(\frac{1 + F_0(x; \theta)}{1 + F(x)}\right) dF(x) = E_F \left[h\left(\frac{1 + F_0(X; \theta)}{1 + F(X)}\right) \right], \tag{2.1}$$

where $E_F[\cdot]$ is the expectation under F and $h : (0, \infty) \rightarrow \mathbb{R}^+$ is assumed to be continuous, decreasing on $(0, 1)$ and increasing on $(1, \infty)$ with an absolute minimum at $x = 1$ such that $h(1) = 0$.

Lemma 2.2. $D(F_0, F) \geq 0$ and equality holds if and only if $F_0 = F$, almost everywhere.

Let us return to the goodness-of-fit test problem for a location-scale family described in (1.1). Firstly, we estimate μ and σ by their maximum likelihood estimators (MLEs), i.e., $\hat{\mu}$ and $\hat{\sigma}$, respectively, and we take $z_i = (x_i - \hat{\mu})/\hat{\sigma}$, $i = 1, \dots, n$. Note that in this family, $F_0(x_i; \hat{\mu}, \hat{\sigma}) = F_0(z_i)$. Secondly, consider the empirical distribution function (EDF) based on data x_i , that is

$$F_n(t) = \frac{1}{n} \sum_{j=1}^n \mathbf{I}_{[x_j \leq t]},$$

where \mathbf{I}_A denotes the indicator of an event A . Then, our proposal is based on the ratio of the standard cdf under H_0 and the EDF based on the x_i 's. Using (2.1) with $F = F_n$, $D(F_0, F_n)$ can be written as

$$H_n := D(F_0, F_n) = \int_{-\infty}^{\infty} h\left(\frac{1 + F_0(x; \hat{\mu}, \hat{\sigma})}{1 + F_n(x)}\right) dF_n(x) = \frac{1}{n} \sum_{i=1}^n h\left(\frac{1 + F_0(z_{(i)})}{1 + i/n}\right)$$

Under H_0 , we expect that $F_0(t; \hat{\mu}, \hat{\sigma}) \approx F_n(t)$, for every $t \in \mathbb{R}$ and $1 + F_0(t; \hat{\mu}, \hat{\sigma}) \approx 1 + F_n(t)$. Note that, since $h(1) = 0$, we expect that $h((1 + F_0(t))/(1 + F_n(t))) \approx 0$ and thus H_n will take values close to zero when H_0 is true. Therefore, it seems justifiable that H_0 must be rejected for large values of H_n . Some standard choices for h are: $h(x) = (x - 1)^2/(x + 1)$, $x \log(x) - x + 1$, $(x - 1) \log(x)$, $|x - 1|$ or $(x - 1)^2$.

Proposition 2.3. *The test statistic based on H_n is invariant under location-scale transformations.*

Proposition 2.4. *Let F_1 be an arbitrary continuous cdf in H_1 . Then under the assumption that the observed sample have cdf F_1 , the test based on H_n is consistent.*

3 A normality test based on H_n

Many statistical procedures are based on the assumption that the observed data are normally distributed. Consequently, a variety of tests have been developed to check the validity of this assumption. In this section, we propose a new normality test based on H_n .

Consider again the goodness-of-fit testing problem described in (1.1), where now $f_0(x; \mu, \sigma) = 1/\sqrt{2\pi\sigma^2}e^{-(x-\mu)^2/2\sigma^2}$, $x \in \mathbb{R}$, in which $\mu \in \mathbb{R}$ and $\sigma > 0$ are both unknown, and $F_0(\cdot; \mu, \sigma)$ is the corresponding cdf, where $F_0(\cdot)$ is the standard case for $F_0(\cdot; 0, 1)$.

First we estimate μ and σ by their maximum likelihood estimators (MLEs), i.e., $\bar{x} = 1/n \sum_{i=1}^n x_i$ and $s^2 = 1/(n-1) \sum_{i=1}^n (x_i - \bar{x})^2$, respectively. Let $z_i = (x_i - \bar{x})/s$, $i = 1, \dots, n$. Then, the test statistic for normality is:

$$H_n = \frac{1}{n} \sum_{i=1}^n h \left(\frac{1 + F_0(x_{(i)}, \bar{x}, s)}{1 + F_n(x_{(i)})} \right) = \frac{1}{n} \sum_{i=1}^n h \left(\frac{1 + F_0(z_{(i)})}{1 + i/n} \right), \quad (3.1)$$

where $h(x) = \left(\frac{x-1}{x+1}\right)^2$. Note that $h : (0, \infty) \rightarrow \mathbb{R}^+$ is decreasing on $(0, 1)$ and increasing on $(1, \infty)$ with an absolute minimum at $x = 1$ such that $h(1) = 0$. We selected this function h , because based on simulation study, it is more powerful than other functions h . For example, we considered $h_2(x) := x \log(x) - x + 1$ for comparison with $h_1(x) := \left(\frac{x-1}{x+1}\right)^2$ (see Table 2).

Table 1 contains the upper critical values of H_n , which have obtained by Monte Carlo from 100000 simulated samples for different sample sizes n and significance levels $\alpha = 0.01, 0.05, 0.1$.

Table 1: Critical values of H_n for $\alpha = 0.01, 0.05, 0.1$.

αn	5	6	7	8	9	10	15	20	25	30	40	50
0.01	.0039	.0035	.0030	.0026	.0023	.0021	.0014	.0011	.0008	.0007	.0005	.0004
0.05	.0030	.0026	.0022	.0019	.0017	.0016	.0010	.0007	.0006	.0005	.0004	.0003
0.10	.0026	.0022	.0019	.0016	.0015	.0013	.0009	.0006	.0005	.0004	.0003	.0002

Remember that, H_n is expected to take values close to zero when H_0 is true. Hence, H_0 will be rejected for large values of H_n . Also H_n is invariant under location-scale transformations and consistent under the assumption H_1 , respectively, from Propositions 2.3 and 2.4.

4 Simulation study

In this section, we study the power of the normality test based on H_n and compare it with a large number of recent and classical normality tests. Following Esteban et al. we consider the following

alternative distributions, that can be classified in four groups:

Group I: Symmetric distributions with support on $(-\infty, \infty)$:

Standard Normal (N); Student's t (t) with 1 and 3 degrees of freedoms; Double Exponential (DE) with parameters $\mu = 0$ (location) and $\sigma = 1$ (scale); Logistic (L) with parameters $\mu = 0$ (location) and $\sigma = 1$ (scale);

Group II: Asymmetric distributions with support on $(-\infty, \infty)$:

Gumbel (Gu) with parameters $\alpha = 0$ (location) and $\beta = 1$ (scale); Skew Normal (SN) with with parameters $\mu = 0$ (location), $\sigma = 1$ (scale) and $\alpha = 2$ (shape);

Group III: Distributions with support on $(0, \infty)$:

Exponential (Exp) with mean 1; Gamma (G) with parameters $\beta = 1$ (scale) and $\alpha = .5, 2$ (shape); Lognormal (LN) with parameters $\mu = 0$ and $\sigma = .5, 1, 2$; Weibull (W) with parameters $\beta = 1$ (scale) and $\alpha = .5, 2$ (shape);

Group IV: Distributions with support on $(0, 1)$:

Uniform (Unif); Beta (B) with parameters (2,2), (.5,.5), (3,1.5) and (2,1).

Table 2 shows the power of the 5% significance level test of normality. Power against an alternative distribution has been estimated by the relative frequency of values of the corresponding statistic in the critical region for 10000 simulated sample of size $n = 10$. The maximum reached power is indicated in bold. In the case of the test based on H_n , we also consider $h_2(x) := x \log(x) - x + 1$ for comparison with $h_1(x) := \left(\frac{x-1}{x+1}\right)^2$.

For testing normality, we consider a large number of recent and classical statistics that have been used to test normality. These tests are:

Vasicek's entropy estimator (KL) [Vasicek \(1976\)](#), Ebrahimi's entropy estimator (TE_{mn}) [Ebrahimi et al. \(1994\)](#), Vasicek's estimator with nonparametric distribution function (TV_{mn}) [Park and Park \(2003\)](#), Ebrahimi estimator with nonparametric distribution function (TE_{mn}) [Park and Park \(2003\)](#), Alizadeh and Arghami estimator with nonparametric distribution function (TA_{mn}) [Alizadeh and Arghami \(2010, 2013\)](#), Dimitriev and Tarasenko's entropy estimator (TD_{mn}) [Dimitriev and arasenko \(1973\)](#), Corea's entropy estimator (TC_{mn}) [Corea \(1995\)](#), Van Es's entropy estimator (TEs_{mn}) [VanEs \(1992\)](#), Zamanzade and Arghami's entropy estimator ($TZ1_{mn}, TZ2_{mn}$) [Zamanzade and Arghami \(2012\)](#), Kolmogorov-Smirnov's (Lilliefors) statistic (kS) [Lilliefors \(1976\)](#), Kuiper's statistic (V) [Kuiper \(1962\)](#), Cramér-von Mises' statistic (W^2), Watson's statistic (U^2) [Watson \(1961\)](#), Anderson-Darling's statistic (A^2) [Anderson \(1954\)](#), Pearson's Chi-square statistic (P) [DAgostino \(1986\)](#), Shapiro-Wilk's statistic (SW) [Shapiro and Wilk \(1965\)](#), Shapiro-Francia's statistic (SF) [Shapiro and Francia \(1972\)](#),

SJ statistic (SJ) [Gel et al. \(2007\)](#), Jarque-Bera's statistic (JB) [Jarque and Bera \(1980\)](#), Robust Jarque-Bera's statistic (RJB) [Gel and Gastwirth \(2008\)](#), Zhang and Wu's statistics (Z_K, Z_C, Z_A) [Zhang and Wu \(2005\)](#), Anscombe and Glynn's statistic (Z_b) [Anscombe and Glynn \(1983\)](#), Bonett and Seier's statistic (Z_w) [Bonett and Seier \(2002\)](#), D'Agostino's statistic (D) [D'Agostino \(1971\)](#), Chen and Shapiro's statistic (QH) [Chen and Shapiro \(1995\)](#), Filliben's statistic (r) [Filliben \(1975\)](#), del Barrio et al.'s statistic (R_n) [Barrio et al. \(1992\)](#), Epps and Pulley's statistic (T_{EP}) [Epps and Pulley \(1983\)](#) and Martinez and Iglewicz's statistic (I_n) [Martinez and Iglewicz \(1981\)](#).

From the simulations results (Table 2) we can conclude that:

- The Robust tests SJ and RJB are the most powerful against alternative distributions that are symmetric with the support on $(-\infty, \infty)$, (Group I).
- The test SF and H_n (the proposed test) are the most powerful against alternative distributions that are asymmetric with the support on $(-\infty, \infty)$, (Group II).
- The tests H_n , QH and TA are the most powerful against alternative distributions with the support on $(0, \infty)$, (Group III).
- The test KL, TV, TC and TA are the most powerful against alternatives with the support $(0, 1)$, (Group IV).

Conclusions

In this paper we propose a statistic to test normality and compared its performance with thirty recent and classical tests for normality. The simulation study reveals that none of the statistics under evaluation can be considered to be the best one for all alternative distributions studied. However, the tests based on RJB or SJ have the best performance in Group I and the same happens to KL or TV in Group IV. Regarding our proposal, H_n is the most powerful in Groups II–III, mainly for small sample sizes.

References

- Vasicek, O.A. (1976). test for normality based on sample entropy, *Journal of the Royal Statistical Society, Series B*, **38**, 54–59.
- Ebrahimi, N., Pflughoeft, K. and Soofi, S.E. (1994). Two measures of sample entropy, *Statistics and Probability Letters*, **20**, 225–234.

- Park, S. and Park, D. (2003). Correcting moments for goodness-of-fit tests based on two entropy estimates, *Journal of Statistical Computation and Simulation*, **73(9)**, 685–694.
- Alizadeh Noughab, H. and Arghami, N.R. (2010). A new estimator of entropy, *Journal of the Iranian Statistical Society*, **9(1)**, 53–64.
- Alizadeh Noughabi H. and Arghami, N.R. (2013). Goodness-of-fit tests based on correcting moments of entropy estimators, *Communications in Statistics - Simulation and Computation*, **42(3)**, 499–513.
- Dimitriev, Y.G. and Tarasenko, F.P. (1973). On the estimation functions of the probability density and its derivatives, *Theory Probab Appl*, **18**, 628–633.
- Corea, J.C. (1995). A new estimator of entropy, *Communications in Statistics -Theory and Methods*, **24**, 2439–2449.
- Van Es, B. (1992). Estimating functionals related to a density by class of statistics based on spacing, *Scand J Statist*, **19**, 61–72.
- Zamanzade, E. and Arghami, N.R. (2012). Testing normality based on new entropy estimators, *Journal of Statistical Computation and Simulation*, **82(11)**, 1701–1713.
- Lilliefors HW. On the kolmogorov-smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association*, **62(318)**, 399–402.
- Kuiper, N.H. (1962). Test concerning random points on a circle, *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, Series A, **63**, 38–47.
- Watson G.S. (1961). Goodness-of-fit tests on a circle, *Technometrics*, **48**, 109–114.
- Anderson, T.W. (1954). A test of goodness of fit, *Journal of the American Statistical Association*, **49**, 765–769.
- D'Agostino, R.B. and Stephens, M. (1986). *Goodness-of-fit techniques*. Marcel Dekker, Inc, New York.
- Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples), *Biometrika*, **52(3/4)**, 591–611.
- Shapiro, S.S. and Francia, R.S. (1972). An Approximate Analysis of Variance Test for Normality, *Journal of the American Statistical Association*, **67(337)**, 215–216.
- Gel, Y.R., Miao, W. and Gastwirth, J.L. (2007). Robust directed tests of normality against heavy-tailed alternatives, *Computational Statistics and Data Analysis*, **51**, 2734–2746.

- Jarque, C.M. and Bera A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, **6**, 255–259.
- Gel Y.R. and Gastwirth, J.L. (2008). A robust modification of the jarque-bera test of normality, *Economics Letters*, **99**, 30–32.
- Zhang, J. and Wu, Y. (2005). Likelihood-ratio tests for normality, *Computational Statistics & Data Analysis*, **49(3)**, 709–721.
- Anscombe, F.J. and Glynn, W.J. (1983). Distribution of kurtosis statistic b_2 for normal statistics. *Biometrika*, **70(1)**, 227-234.
- Bonett D.G. and Seier, E. (2002). A test of normality with high uniform power. *Computational Statistics & Data Analysis*, **40(3)**, 435-445.
- D'Agostino R.B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, **58(2)**, 341–348.
- Chen, L. and Shapiro, S.S. (1995). An alternative test for normality based on normalized spacings, *Journal of Statistical Computation and Simulation*, **53(3-4)**, 269–287.
- Filliben, J.J. (1975). The probability plot correlation coefficient test for normality, *Technometrics*, **17(1)**, 111–117.
- Barrio, E.d., Cuesta-Albertos, J.A. and Matrn, C. and Rodrguez-Rodrguez, J.M. Tests of goodness of fit based on the l2-wasserstein distance, *The Annals of Statistics*, **27(4)**, 1230-1239.
- Epps, T.W. and Pulley, L.B. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, **70(3)**, 723-726.
- Martinez, J. and Iglewicz, B. (1981). A test for departure from normality based on a biweight estimator of scale, *Biometrika*, **68(1)**, 331–333.

Table 2: Power comparisons for the normality test, $\alpha = 0.05$, $n = 10$.

Group altern.	N	I				II		III							IV					
		t(1)	t(3)	L	DE	Gu	SN	Exp	G(2)	G(.5)	LN(1)	LN(2)	LN(.5)	W(.5)	W(2)	Unif	B(2,2)	B(.5,.5)	B(3,.5)	B(2,1)
KL	.048	.442	.091	.051	.091	.101	.058	.416	.179	.782	.552	.938	.181	.931	.075	.167	.082	.512	.108	.173
TD	.051	.583	.201	.087	.163	.154	.071	.394	.222	.631	.565	.869	.249	.813	.076	.028	.025	.080	.065	.093
TEs	.049	.591	.167	.074	.140	.113	.062	.330	.158	.679	.485	.892	.176	.876	.064	.061	.037	.238	.064	.092
TC	.054	.409	.083	.047	.057	.097	.053	.404	.173	.786	.542	.936	.171	.926	.071	.170	.086	.489	.110	.182
TZ1	.053	.632	.212	.089	.177	.145	.068	.359	.209	.581	.524	.846	.229	.784	.074	.030	.025	.078	.061	.081
TZ2	.051	.638	.216	.091	.181	.144	.066	.353	.205	.572	.516	.840	.228	.776	.073	.026	.023	.060	.058	.076
TV	.048	.375	.082	.048	.053	.092	.055	.397	.151	.762	.519	.933	.144	.923	.073	.181	.084	.514	.656	.170
TE	.052	.460	.112	.058	.077	.111	.059	.454	.185	.794	.581	.945	.181	.935	.074	.158	.071	.481	.686	.164
TA	.053	.507	.134	.065	.094	.124	.062	.477	.213	.810	.616	.951	.208	.940	.080	.129	.064	.451	.704	.162
ZK	.051	.583	.173	.078	.140	.116	.067	.337	.178	.635	.502	.878	.189	.848	.075	.077	.047	.205	.490	.103
ZC	.044	.562	.183	.079	.142	.144	.070	.437	.236	.731	.598	.918	.241	.897	.083	.083	.038	.316	.598	.125
ZA	.045	.594	.197	.086	.160	.149	.073	.445	.245	.738	.610	.924	.250	.903	.081	.046	.028	.196	.605	.111
Zb	.048	.520	.169	.071	.131	.107	.062	.218	.149	.343	.350	.600	.162	.499	.070	.117	.059	.277	.231	.092
Zw	.050	.501	.162	.073	.138	.077	.058	.126	.091	.187	.222	.424	.103	.313	.057	.103	.060	.228	.129	.075
D	.052	.584	.178	.076	.142	.111	.068	.263	.153	.495	.437	.803	.178	.725	.064	.042	.044	.037	.332	.057
QH	.046	.584	.186	.081	.147	.147	.072	.439	.241	.737	.605	.923	.244	.901	.083	.084	.041	.302	.604	.127
r	.045	.623	.210	.090	.176	.147	.076	.411	.228	.685	.580	.903	.243	.872	.080	.043	.027	.158	.540	.098
Rn	.044	.595	.190	.083	.156	.147	.071	.432	.240	.726	.599	.916	.245	.895	.082	.072	.036	.258	.589	.118
T _{EP}	.046	.587	.195	.086	.157	.151	.076	.411	.239	.663	.583	.884	.250	.844	.085	.054	.033	.143	.511	.111
I _n	.047	.159	.144	.086	.147	.111	.068	.212	.147	.210	.213	.102	.158	.134	.070	.027	.027	.046	.196	.066
P	.042	.531	.148	.083	.136	.127	.080	.397	.200	.704	.545	.903	.199	.878	.087	.086	.061	.229	.594	.136
KS	.051	.580	.164	.073	.142	.121	.065	.301	.175	.540	.463	.826	.182	.758	.074	.066	.046	.162	.418	.100
V	.054	.589	.163	.071	.142	.117	.060	.360	.180	.662	.524	.892	.187	.854	.068	.081	.048	.237	.530	.109
W ²	.047	.618	.182	.080	.158	.137	.066	.390	.210	.672	.554	.896	.220	.855	.079	.074	.044	.229	.542	.115
U ²	.050	.604	.175	.074	.150	.121	.067	.366	.196	.655	.535	.885	.202	.856	.074	.082	.047	.243	.510	.113
A ²	.048	.618	.190	.083	.159	.147	.068	.416	.225	.703	.578	.909	.233	.875	.083	.080	.046	.268	.576	.126
SW	.046	.594	.187	.082	.150	.153	.071	.442	.239	.735	.603	.920	.245	.894	.084	.082	.042	.299	.609	.130
SF	.048	.631	.214	.098	.193	.161	.078	.432	.242	.706	.600	.912	.259	.877	.087	.050	.030	.184	.576	.108
SJ	.056	.680	.257	.124	.258	.149	.086	.291	.182	.469	.456	.785	.209	.696	.081	.021	.031	.026	.317	.056
JB	.054	.590	.212	.094	.175	.160	.071	.371	.225	.598	.549	.848	.241	.784	.079	.028	.022	.096	.445	.083
RJB	.056	.644	.235	.106	.213	.150	.074	.320	.199	.504	.484	.792	.224	.708	.077	.019	.020	.026	.348	.056
new test																				
H _n , h ₂	.048	.596	.176	.071	.149	.189	.089	.504	.285	.774	.654	.940	.290	.917	.107	.074	.046	.212	.311	.051
H _n , h ₁	.052	.587	.173	.070	.146	.198	.093	.516	.296	.781	.662	.942	.301	.919	.113	.079	.049	.216	.285	.047



Prediction For Lindley Distribution Based on Progressively Type-II Censored Samples

R. Valiollahi¹, A. Asgharzadeh²

¹Department of Statistics, Semnan University, Semnan, Iran

²Department of Statistics, University of Mazandaran, Babolsar, Iran

Abstract: Prediction of unobserved or censored observations is an interesting topic in reliability and life-testing experiments. This paper considers the prediction problem of times to failure of units censored in multiple stages in a progressively censored sample from the Lindley distribution. The maximum likelihood predictor, the best linear predictor and the conditional median predictor are provided. A numerical example and a Monte Carlo simulation study are given to illustrate the prediction methods discussed in this paper.

Keywords: Lindley Distribution, Progressively Type II censored sample, Maximum likelihood prediction, Best unbiased prediction, Conditional median prediction.

Mathematics Subject Classification (2010): 99X99 99X99 99X99.

1 Introduction

In the recent years, the Lindley distribution has received a considerable attention in the statistical literature in the recent years. The Lindley distribution has the probability density function (pdf)

$$f(x; \theta) = \frac{\theta^2}{1 + \theta}(1 + x)e^{-\theta x}, \quad x > 0, \quad \theta > 0. \quad (1.1)$$

and the corresponding cumulative distribution function (cdf) given by

$$F(x; \theta) = 1 - \frac{1 + \theta + \theta x}{1 + \theta}e^{-\theta x}, \quad x > 0, \quad \theta > 0. \quad (1.2)$$

The Lindley distribution was originally introduced by Lindley (1958) in the context of Bayesian statistics. Recently, this distribution has been studied and generalized by several authors, see Zakerzadeh and Dolati (2009), and Ghitany et al. (2013). Estimation for the Lindley distribution has been discussed by many authors. See for example Krishna and Kumar (2011), Gupta and Singh (2013) and Al-Mutairi et al. (2013).

²R. Valiollahi : rvaliollahi.reza@semnan.ac.ir

Prediction of unobserved or censored observations based on censored samples has been discussed by many authors . For some recent references, see [Raqab et al. \(2010\)](#) and [Agharzadeh et al. \(2013\)](#). The aim of this paper is to consider the prediction problem for the Lindley distribution based on progressive Type-II censored samples. We study this problem via non-Bayesian approach and provide several predictors of future failures. To the best of our knowledge, this problem has not been studied before in the literature.

The progressive Type-II censoring can be described as follows. Suppose n units are placed on a life test. Immediately following the first failure, R_1 surviving units are removed from the test at random. Then, immediately following the second failure, R_2 surviving units are removed from the test at random. This process continues until, at the time of the $m - th$ failure, all the remaining $R_m = n - R_1 - R_2 - \dots - R_{m-1} - m$ units are removed from the experiment. Here the R_i s are fixed prior to study. If $R_1 = R_2 = \dots = R_m = 0$, then $n = m$ which corresponds to the complete sample situation. If $R_1 = R_2 = \dots = R_{m-1} = 0$, we have $R_m = n - m$ which corresponds to the conventional Type-II right censoring scheme. For more details on progressively censoring, inferences and their applications, one may refer to [Balakrishnan and Aggarwala \(2000\)](#).

2 Log-Likelihood Function

Let $X_{1:m:n}, \dots, X_{m:m:n}$ denote a progressively Type-II censored sample from the $LD(\theta)$, obtained from a sample of size n with the censoring scheme $R = (R_1, \dots, R_m)$. To simplify the notation, we will use X_i in place of $X_{i:m:n}$. On the basis of $\mathbf{X} = (X_1, X_2, \dots, X_m)$, the likelihood function is given (see [Balakrishnan and Aggarwala \(2000\)](#)) by

$$\begin{aligned}
 L(\theta) &= A \prod_{i=1}^m \left\{ \frac{\theta^2}{1+\theta} (1+x_i) e^{-\theta x_i} \right\} \left(\frac{1+\theta+\theta x_i}{1+\theta} e^{-\theta x_i} \right)^{R_i} \\
 &= A \frac{\theta^{2m}}{(1+\theta)^n} \prod_{i=1}^m \left\{ (1+x_i)(1+\theta+\theta x_i)^{R_i} \right\} e^{-\theta \sum_{i=1}^m (R_i+1)x_i},
 \end{aligned}
 \tag{2.1}$$

where $A = n(n-1-R_1)(n-2-R_1-R_2) \dots (n-m+1-R_1 \dots -R_{m-1})$.

The log-likelihood function is

$$\ln L(\theta) \propto 2m \ln \theta - n \ln(1+\theta) + \sum_{i=1}^m R_i \ln(1+\theta+\theta x_i) - \theta \sum_{i=1}^m (R_i+1)x_i.
 \tag{2.2}$$

From the log-likelihood function, we obtain the log-likelihood equation as

$$\frac{d \ln L(\theta)}{d\theta} = \frac{2m}{\theta} - \frac{n}{1+\theta} + \sum_{i=1}^m R_i \frac{(1+x_i)}{1+\theta+\theta x_i} - \sum_{i=1}^m (R_i+1)x_i = 0.
 \tag{2.3}$$

So, the maximum likelihood estimate (MLE) of θ can be obtained by solving the log-likelihood equation using some numerical methods.

3 Point Prediction

In this section, we provide several point predictors of $Y = X_{s:R_i}$, ($s = 1, \dots, R_i$, $i = 1, \dots, m$) based on the observed censored sample \mathbf{X} . The conditional distribution of $Y = X_{s:R_i}$ given \mathbf{X} is just the distribution of $Y = X_{s:R_i}$ given $X_i = x_i$ (due to the Markovian property of progressively Type-II right censored-order statistics). Consequently, the density of $Y = X_{s:R_i}$ given $\mathbf{X} = \mathbf{x}$ is the same as the density of the s th order statistic out of R_i units from the population with density $f(y)/(1 - F(x_i))$, $y \geq x_i$ (left truncated density at x_i). So, the conditional density of Y given $\mathbf{X} = \mathbf{x}$, for $y \geq x_i$, is given by

$$f_{Y|X_i}(y|x_i, \theta) = s \binom{R_i}{s} f(y) [F(y) - F(x_i)]^{s-1} [1 - F(y)]^{R_i-s} [1 - F(x_i)]^{-R_i}. \quad (3.1)$$

For Lindely model, (3.1) reduces to

$$f(y|\mathbf{x}, \theta) = s \binom{R_i}{s} \theta^2 (1+y)(1+\theta+\theta y)^{R_i-s} (1+\theta+\theta x_i)^{-R_i} \\ \times e^{-\theta(R_i-s+1)y+\theta R_i x_i} [(1+\theta+\theta x_i)e^{-\theta x_i} - (1+\theta+\theta y)e^{-\theta y}]^{s-1}. \quad (3.2)$$

3.1 Maximum Likelihood Predictor

The principle of maximum likelihood is used in maximum likelihood prediction method to the joint prediction and estimation of a future random variable and an unknown parameter. The predictive likelihood function (PLF) of Y and θ is

$$L(y, \theta|\mathbf{x}) = f(y|\mathbf{x}, \theta) f(\mathbf{x}|\theta). \quad (3.3)$$

If $\hat{Y} = u(\mathbf{X})$ and $\hat{\theta} = \nu(\mathbf{X})$ are statistics for which

$$L(u(\mathbf{x}), \nu(\mathbf{x})|\mathbf{x}) = \sup_{(y, \theta)} L(y, \theta|\mathbf{x}), \quad (3.4)$$

then $u(\mathbf{X})$ is said to be the maximum likelihood predictor (MLP) of the Y and $\nu(\mathbf{X})$ is the predictive maximum likelihood estimator (PMLE) of θ respectively. Consequently, for the Lindley model, the PLF of Y and θ can be obtain as

$$L(y, \theta) = s \binom{R_i}{s} \frac{\theta^{2m+2}}{(1+\theta)^n} (1+y)(1+x_i) \prod_{j=1, j \neq i}^m (1+x_j)(1+\theta+\theta x_j)^{R_j}$$

$$\begin{aligned} & \times (1 + \theta + \theta y)^{R_i - s} \exp \left\{ -\theta \left((R_i - s + 1)y + \sum_{j=1, j \neq i}^m (R_j + 1)x_j - x_i \right) \right\} \\ & \times ((1 + \theta + \theta x_i)e^{-\theta x_i} - (1 + \theta + \theta y)e^{-\theta y})^{s-1}. \end{aligned} \tag{3.5}$$

Apart from a constant term, the predictive log-likelihood function is

$$\begin{aligned} \ln L(y, \theta) & \propto (2m + 2) \ln \theta - n \ln(1 + \theta) + \ln(1 + y) + \sum_{j=1, j \neq i}^m R_j \ln(1 + \theta + \theta x_j) \\ & + (R_i - s) \ln(1 + \theta + \theta y) - \theta \left((R_i - s + 1)y + \sum_{j=1, j \neq i}^m (R_j + 1)x_j - x_i \right) \\ & + (s - 1) \ln((1 + \theta + \theta x_i)e^{-\theta x_i} - (1 + \theta + \theta y)e^{-\theta y}). \end{aligned} \tag{3.6}$$

By using (3.6), the predictive likelihood equations(PLEs) for y and θ are obtained as

$$\begin{aligned} \frac{\partial \ln L(y, \theta)}{\partial y} & = \frac{1}{y + 1} - \theta(R_i - s + 1) + (R_i - s) \frac{\theta}{1 + \theta + \theta y} \\ & + (s - 1) \frac{\theta^2(1 + y)e^{-\theta y}}{(1 + \theta + \theta x_i)e^{-\theta x_i} - (1 + \theta + \theta y)e^{-\theta y}}, \end{aligned} \tag{3.7}$$

$$\begin{aligned} \frac{\partial \ln L(y, \theta)}{\partial \theta} & = \frac{2m + 2}{\theta} - \frac{n}{1 + \theta} + \sum_{j=1, j \neq i}^m R_j \frac{1 + x_j}{1 + \theta + \theta x_j} \\ & + (R_i - s) \frac{1 + y}{1 + \theta + \theta y} - \left((R_i - s + 1)y + \sum_{j=1, j \neq i}^m (R_j + 1)x_j - x_i \right) \\ & + (s - 1) \frac{e^{-\theta x_i}(1 - \theta x_i(1 + x_i)) - e^{-\theta y}(1 - \theta y(1 + y))}{(1 + \theta + \theta x_i)e^{-\theta x_i} - (1 + \theta + \theta y)e^{-\theta y}}. \end{aligned} \tag{3.8}$$

By solving (3.7) and (3.8) with respect to y and θ simultaneously, the MLP of $Y = X_{s:R_i}$, \hat{Y}_{MLP} , and PMLE of θ can be obtained. Some numerical methods are needed to solve this equation and obtain the MLP \hat{Y}_{MLP} and PMLE of θ .

3.2 Best Unbiased Predictor

Let the statistic \hat{Y} is used to predict $Y = X_{s:R_i}$. The predictor \hat{Y} is called a best unbiased predictor (BUP) of Y , if the predictor error $\hat{Y} - Y$ has a mean zero and its prediction error variance $\text{Var}(\hat{Y} - Y)$

is less than or equal to that of any other unbiased predictor of Y . The BUP of Y is

$$\widehat{Y}_{BUP} = E(Y|\mathbf{X}) = \int_{x_i}^{\infty} yf(y|\mathbf{x})dy. \quad (3.9)$$

By using

$$\begin{aligned} & [(1 + \theta + \theta x_i)^{-\theta x_i} - (1 + \theta + \theta y)^{-\theta y}]^{s-1} = \\ & \sum_{j=0}^{s-1} \binom{s-1}{j} (-1)^{(s-j-1)} (1 + \theta + \theta x_i)^j (1 + \theta + \theta y)^{s-j-1} e^{-\theta j x_i} e^{-\theta(s-j-1)y}, \end{aligned} \quad (3.10)$$

we obtain

$$\begin{aligned} \widehat{Y}_{BUP} &= s \binom{R_i}{s} \theta^2 \sum_{j=0}^{s-1} \binom{s-1}{j} (-1)^{(s-j-1)} (1 + \theta + \theta x_i)^{-(R_i-j)} e^{\theta(R_i-j)x_i} \\ &\times \int_{x_i}^{\infty} y(1+y)(1 + \theta + \theta y)^{R_i-j-1} e^{-\theta(R_i-j)y} dy \end{aligned} \quad (3.11)$$

Because the parameter θ is unknown, it has to be estimated. Thus one would replace it by its MLE and obtain the BUP of Y .

3.3 Conditional Median Predictors

Conditional median predictor (CMP) was first proposed by [Raqab and Nagaraja \(1995\)](#). A predictor \widehat{Y} is called the CMP of Y , if it is the median of the conditional distribution of Y given $X_i = x_i$ that is

$$P_{\theta}(Y \leq \widehat{Y}|\mathbf{X} = \mathbf{x}) = P_{\theta}(Y \geq \widehat{Y}|\mathbf{X} = \mathbf{x}). \quad (3.12)$$

We have

$$\begin{aligned} & P_{\theta}(Y \leq \widehat{Y}|\mathbf{X} = \mathbf{x}) = \\ & P_{\theta}\left(1 - \frac{(1 + \theta + \theta Y)e^{-\theta Y}}{(1 + \theta + \theta X_i)e^{-\theta X_i}} \leq 1 - \frac{(1 + \theta + \theta \widehat{Y})e^{-\theta \widehat{Y}}}{(1 + \theta + \theta X_i)e^{-\theta X_i}} \mid \mathbf{X} = \mathbf{x}\right). \end{aligned} \quad (3.13)$$

Now, using the fact that the distribution of $1 - \frac{(1 + \theta + \theta Y)e^{-\theta Y}}{(1 + \theta + \theta X_i)e^{-\theta X_i}}$ given $\mathbf{X} = \mathbf{x}$ is a $Beta(s, R_i - s + 1)$ distribution, the CMP of Y can be obtained by solving the equation

$$1 - \frac{(1 + \theta + \theta \widehat{Y})e^{-\theta \widehat{Y}}}{(1 + \theta + \theta X_i)e^{-\theta X_i}} = Med(B), \quad (3.14)$$

where B has $Beta(s, R_i - s + 1)$ distribution and $Med(B)$ stands for median of B . By doing some calculations, we obtain

$$\hat{Y}_{CMP} = -1 - \frac{1}{\theta} - \frac{1}{\theta} W_{-1} \left([Med(B) - 1](1 + \theta + \theta X_i)e^{-(1+\theta+\theta X_i)} \right) \tag{3.15}$$

where $W_{-1}(\cdot)$ denotes the negative branch of the Lambert W function (i.e., the solution of the equation $W(z)e^{W(z)} = z$). By substituting θ with its MLE, we obtain the CMP of Y .

4 Real Data Analysis and Simulation

To illustrate the prediction methods developed in this paper, we present the analysis of a real data set and simulation study here.

5 Real Data Analysis

The following data set presented in [Murthy et al. \(2004\)](#). The data shows the time between failures for repairable items.

0.11	0.30	0.40	0.45	0.59	0.63	0.70	0.71	0.74	0.77
0.94	1.06	1.17	1.23	1.23	1.24	1.43	1.46	1.49	1.74
1.82	1.86	1.97	2.23	2.37	2.46	2.63	3.46	4.36	4.73

To check the validity of the use of Lindley distribution to fit this data set, Kolmogorov-Smirnov (K-S) test is applied. The the K-S statistics of the distance between the fitted and the empirical distribution functions (based on the parameter $\theta = 0.976$) is 0.561 and the corresponding p -value is 0.141. Therefore, it is reasonable to use the Lindley distribution to fit the data.

We consider the progressively censored scheme $R = (0, 0, 4, 0, 0, 6, 0, 0, 0, 10)$, so the vector of observed failure times is $\mathbf{X} = (0.11, 0.30, 0.40, 0.71, 0.74, 0.77, 1.43, 1.46, 1.49, 1.74)$. Using this sample, we obtain the point predictors MLP, BUP and CMP for $Y = X_{s:R_i}$, ($s = 1, 2, 3; R_i = 3, 6, 10$). The results are displayed in Table 1.

Table 1: Point and interval predictors for Y.

	$i = 3$			$i = 6$			$i = 10$		
	MLP	BUP	CMP	MLP	BUP	CMP	MLP	BUP	CMP
s=1	0.400	1.154	0.967	0.770	1.253	1.122	1.740	1.996	1.921
s=2	1.353	2.035	1.843	1.317	1.788	1.657	2.001	2.273	2.195
s=3	2.385	3.220	2.984	1.904	2.411	2.274	2.282	2.257	2.496

Now, a Monte Carlo simulation study is used to evaluate the biases and MSPEs for the predictors MLP, BUP and CMP. We randomly generated 2000 progressively censored sample from Lindley distribution with $\theta = 0.75, 1.5$. We also used $R = (0, 0, 3, 0, 3, 0, 0, 5)$ as a censoring scheme. Table 2 displays the biases and MSPEs of different predictors obtained from this simulation study.

Table 2: Biases and MSPEs of point predictors.

		$\theta = 0.75$			$\theta = 1.5$			
		MLP	BUP	CMP	MLP	BUP	CMP	
	s=1	Bias	-0.244	-0.105	-0.194	-0.321	-0.073	-0.087
		MSPE	0.615	0.459	0.497	0.195	0.092	0.099
$i = 3$	s=2	Bias	-0.307	-0.142	-0.239	-0.380	-0.092	-0.098
		MSPE	0.899	0.624	0.696	0.437	0.270	0.279
	s=3	Bias	-0.439	-0.306	-0.340	-0.465	-0.176	-0.190
		MSPE	1.736	1.244	1.386	1.436	0.912	0.948
	s=1	Bias	-0.352	-0.184	-0.202	-0.310	-0.065	-0.076
		MSPE	0.704	0.473	0.513	0.212	0.109	0.127
$i = 5$	s=2	Bias	-0.599	-0.205	-0.237	-0.389	-0.105	-0.115
		MSPE	0.959	0.723	0.879	0.463	0.294	0.314
	s=3	Bias	-0.632	-0.302	-0.381	-0.644	-0.205	-0.228
		MSPE	2.331	1.704	1.853	1.503	0.985	1.021
	s=1	Bias	-0.274	-0.152	-0.191	-0.321	-0.074	-0.087
		MSPE	0.794	0.444	0.481	0.195	0.092	0.099
	s=2	Bias	-0.387	-0.209	-0.225	-0.392	-0.110	-0.135
		MSPE	0.943	0.625	0.676	0.452	0.277	0.287
$i = 8$	s=3	Bias	-0.523	-0.327	-0.365	-0.643	-0.202	-0.219
		MSPE	1.802	0.862	0.893	1.406	0.887	0.921
	s=4	Bias	-0.603	-0.390	-0.400	-0.670	-0.263	-0.266
		MSPE	2.421	1.273	1.352	1.839	1.154	1.218
	s=5	Bias	-0.927	-0.529	-0.552	-0.830	-0.441	-0.472
		MSPE	2.946	1.839	2.004	2.090	1.343	1.400

Conclusion

From Table 2, we observe that the BUP is the best predictor in terms of biases and MSPEs. The CMP is the second best predictor. Also, the MLP does not work well.

References

- A. Asgharzadeh, R. Valiollahi, D. Kundu. (2013). Prediction for future failures in Weibull distribution under hybrid censoring. Journal of Statistical Computation and Simulation, <http://dx.doi.org/10.1080/00949655.2013.848451>

- Al-Mutairi, D.K., Ghitany, M. E., and Kundu, D. (2013). Inferences on Stress-Strength Reliability from Lindley Distributions. *Communications in Statistics - Theory and Methods*, **42(8)**.
- Balakrishnan, N. and Aggarwala, R. (2000). *Progressive censoring: Theory, Methods and Applications*. Boston: Birkhäuser.
- Ghitany, M.E., Al-Mutairi, D.K., Balakrishnan, N., Al-Enezi, L.J. (2013). Power Lindley distribution and associated inference. *Computational Statistics and Data Analysis*, **64**, 20–33.
- Gupta, P. K. and Singh. B. (2013). Parameter estimation of Lindley distribution with hybrid censored data. *International Journal of System Assurance Engineering and Management*, **4(4)**, 378–385.
- Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society*, **A**, **20**, 102–107.
- Krishna, H., and Kumar, K. (2011). Reliability estimation in Lindley distribution with progressively type II right censored sample. *Mathematics and Computers in Simulation*, **82**, 281–294.
- Zakerzadeh, H., Dolati, A. (2009). Generalized Lindley distribution. *Journal of Mathematical Extension*, **3**, 13–25.
- M. Z. Raqab, A. Asgharzadeh, and R. Valiollahi, (2010). Prediction for Pareto distribution based on progressively Type-II censored samples. *Computational Statistics and Data Analysis*, **54**, 1732-1743.
- Raqab, M. Z. and Nagaraja, H. N. (1995). "On some predictors of future order statistics," *Metro*, **53(12)**, 185–204.
- Murthy, D. N. P. Xie, M. and Jiang, R. (2004). "Weibull Models", *Wiley series in probability and statistics*, John Wiley and Sons, NJ.



A new generalized Pareto - Poisson distribution generated by beta random variables

Shahram Yaghoobzadeh Shahrastani¹
University of Payame Noor, Soomehsara, Iran

Abstract: In this paper, we introduce a new five-parameters distribution called as the beta Paretopoisson (BPP) distribution. We obtain several properties of the new distribution such as its probability density function, its Moments of order statistics and Mean deviations. The maximum likelihood estimation procedure is presented in this paper, then usefulness of distribution (BPP) is shown using a real dataset.

Keywords: Pareto distribution, beta Paretopoisson distribution, Maximum likelihood estimation, beta distribution, Order Statistics.

Mathematics Subject Classification (2010): 62E10.

1 Introduction

The family of Pareto distributions is well known in the literature for its capability in modelling the heavy-tailed distributions that are mostly common in data on income distribution, city population size, and many other quantities measured in the physics, biology, hydrology, and engineering. Generalized distributions have been widely studied in statistics as they possess flexibility in applications. This is what urged, [Gupta and Kundu \(2001\)](#), [Nadarajah \(2005\)](#) and [Cooray \(2010\)](#) to the introduce exponentiated distributions and derive several of their properties. The proposed generalization stems from the following general class: if G denotes the cumulative distribution function (cdf) of a random variable then a generalized class of distributions can be defined by

$$F(x) = I_{G(x)}(a, b) \quad (1.1)$$

For $a > 0$ and $b > 0$, where $I_y(a, b) = B_y(a, b)/B(a, b)$ denotes the incomplete beta function ratio, and

$$B_y(a, b) = \int_0^y u^{a-1}(1-u)^{b-1} du$$

denotes the incomplete beta function. In recent years, several new distributions have been proposed by authors using Equation (1.1) in the literature. [Akinsete et al. \(2008\)](#) introduce the four parameter

¹Speaker's Name: Shahram Yaghoobzadeh Shahrastani yagoubzade@gmail.com

beta Pareto (BP) distribution following the work proposed by Eugene et al. (2002). Nassar and Nada (2011) proposed a five parameter distribution, the so-called beta generalized Pareto (BGP) distribution. The beta exponential (BE) distributions is proposed by Nadarajah and Kotz (2006), which $G(x)$ is the cdf of the exponential distribution in the second. the beta modified Weibull (BMW) distribution introduced by Silva et al. (2010), and the beta Weibull-geometric (BWG) distribution introduced by Cordeiro et al. (2011) which extends the WG distribution of Barreto-Souza et al. (2011). The paper is organized as follows. In Section 2, we review the Pareto Possion (PP) distribution (Martines and Stedinger (2001)) and define the BPP distribution. The density functions, the moment generating function (mgf), the moments, the moments of order statistics and Mean deviations from the mean and median are obtained in this Section. Maximum-likelihood estimation of the model parameters and the observed information matrix are discussed in Section 3. In Section 4, the flexibility and potentiality of the new distribution are illustrated using a real data set and the new model will be compared with some sub-models and Some concluding remarks are given in Section 5. The density function of a *Beta – G* distribution corresponding to cdf (1.1) is given by

$$f(x, a, b) = \frac{g(x)}{B(a, b)} (G(x))^{a-1} (1 - G(x))^{b-1} \tag{1.2}$$

where $g(x)$ is the density function corresponding to cdf $G(x)$.

2 BPP distribution

The random variable X has an PP distribution if its cumulative distribution function (cdf) takes the form

$$G(x, \gamma, \theta, \mu) = \frac{e^\theta - e^{\theta(\frac{x}{\mu})^\gamma}}{e^\theta - 1}, \quad x > \mu > 0 \tag{2.1}$$

where $\gamma > 0, \theta > 0$. The corresponding probability density function (pdf) is

$$g(x, \gamma, \theta, \mu) = \frac{\gamma \theta \mu^\gamma e^{\theta(\frac{x}{\mu})^\gamma}}{x^{\gamma+1} (e^\theta - 1)}, \quad x > \mu > 0, \quad \theta > 0 \tag{2.2}$$

Using the $u = \theta(\frac{x}{\mu})^\gamma$, The k th moment about zero of the PP distribution is given by

$$E(X^k) = \frac{\theta^{\frac{1}{\gamma}} \mu \varphi(1 - \frac{1}{\gamma}, \theta)}{e^\theta - 1} \tag{2.3}$$

where $\varphi(r, s) = \int_0^s u^{r-1} e^u du$. Replacing $G(x)$ in Equation (1.1) by the (pdf) (2.1) yields the new cdf

$$F(x, a, b, \gamma, \theta, \mu) = \frac{1}{B(a, b)} \int_0^{(e^\theta - e^{\theta(\frac{x}{\mu})^\gamma})/(e^\theta - 1)} u^{a-1} (1 - u)^{b-1} du, \quad x > 0 \tag{2.4}$$

where $\gamma > 0, \mu > 0$. A random variable X with the cdf (2.4) is said to have a BPP distribution and will be denoted by $X \sim BPP(a, b, \gamma, \theta, \mu)$. From Equations (1.2) and (2.1), the pdf, the cdf of a BPP distribution are given, respectively, by

$$f(x, a, b, \gamma, \theta, \mu) = \frac{\gamma\theta\mu^\gamma e^{\theta(\frac{\mu}{x})^\gamma} (e^\theta - e^{\theta(\frac{\mu}{x})^\gamma})^{a-1} (e^{\theta(\frac{\mu}{x})^\gamma} - 1)^{b-1}}{x^{\gamma+1} B(a, b) (e^\theta - 1)^{a+b-1}}, \quad x > 0 \tag{2.5}$$

and

$$F(x, a, b, \gamma, \theta, \mu) = \frac{1}{B(a, b)} \sum_{k=0}^{\infty} \binom{b}{k} (-1)^k \left(\frac{e^\theta - e^{\theta(\frac{\mu}{x})^\gamma}}{e^\theta - 1}\right)^{a+k} \frac{1}{a+k} \tag{2.6}$$

where $I_{\{1-G(x,\gamma,\theta,\mu)\}^{(b,a)}} = 1 - \bar{F}(x, a, b, \gamma, \theta, \mu)$.

Figure 1 plot the failure-rate function of the BPP distribution for some choices of the parameter values. As we see from Figure 1, a characteristic of the BPP distribution is that its failure-rate function can be decreasing, increasing, bathtub, or unimodal depending on its parameter values.

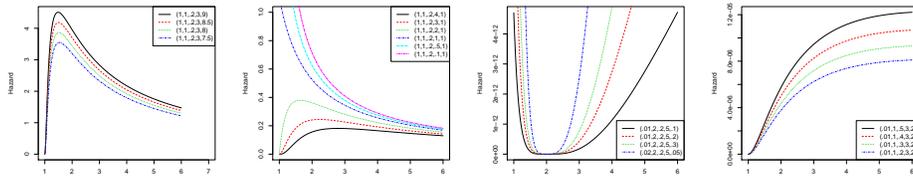


Figure 1: Plots of hazard rate function of BPP for different values of parameters $(\gamma, \mu, \theta, a, b)$.

2.1 Mgf and Moments of BPP

Here, we first present an alternative expression for the pdf of a BPP distribution. Using the power series expansion $(1 - z)^\alpha = \sum_{j=0}^{\infty} (-1)^j \binom{\alpha}{j} z^j$ in Equation (2.6) implies for real non-integer $a > 0$ is

$$f(x, a, b, \gamma, \theta, \mu) = \sum_{i=0}^{\infty} w_{i,a,b,\theta} f_{pp}(\gamma, \theta_j, \mu) \tag{2.7}$$

where $\theta_{i,j} = \theta(j + 1)$ and

$$w_{i,a,b,\theta} = \frac{1}{(e^\theta - 1)^{a+b-2}} \sum_{j=0}^{\infty} \frac{\Gamma(a+i)e^{\theta(a+i-j-1)} (-1)^{i+j} \Gamma(a+b)}{i! \Gamma(j) \Gamma(b-i) \Gamma(a+i-j) (e^\theta - 1)^{a+i-1}}$$

Now we obtain the moment generating function of the BPP distribution.

Suppose that $Y \sim BPP(a, b, \gamma, \theta, \mu)$ and $X \sim PP(\gamma, \theta, \mu)$. According to (2.3), We obtain,

$$M_X(t) = \frac{1}{e^\theta - 1} \sum_{r=0}^{\infty} \frac{(t\mu\theta^{\frac{1}{\gamma}})^r \varphi(\frac{\gamma-r}{\gamma}, \theta)}{r!} \tag{2.8}$$

Hence, combining Equations (2.5) and (2.8), we obtain the mgf of the BPP distribution

$$M_Y(t) = \frac{1}{e^\theta - 1} \sum_{i=0}^{\infty} \sum_{r=0}^{\infty} \frac{w_{i,a,b,\theta} (t\mu(\theta_j)^{\frac{1}{\gamma}})^r \varphi(\frac{\gamma-r}{\gamma}, \theta_j)}{r!} \tag{2.9}$$

According to (2.3) and (2.5), We obtain the moments of the BPP distribution

$$E(Y^r) = \frac{1}{e^\theta - 1} \sum_{i=0}^{\infty} \frac{w_{i,a,b,\theta} (t\mu(\theta_j)^{\frac{1}{\gamma}})^r \varphi(\frac{\gamma-r}{\gamma}, \theta_j)}{r!}. \tag{2.10}$$

2.2 Moments of order statistics

Cordeiro et al. (2011), showed that the density function of the i ,th order statistic $X_{i:n}$, say $f_{i:n}$, For a (BG) model defined from the parent functions $g(x)$ and $G(x)$ for real non-integer $a > 0$ is

$$f_{i:n}(x) = \frac{g(x)(G(x))^{a-1}(1-G(x))^{b-1}}{B(a,b)B(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \sum_{m=0}^{\infty} c_{i+j-1,m} (G(x))^m \tag{2.11}$$

where the coefficients $c_{i+j-1,m}$ follow from Equation $c_{j,i} = (ia_0)^{-1} \sum_{m=1}^i (jm - i + m)a_m c_{j,i-m}$ and $c_{j,0} = a_0^j$. Hence, $c_{j,i}$ can be calculated from $c_{j,1}, \dots, c_{j,i-1}$ and then from a_0, \dots, a_i . where $a_m = \frac{(1-b)_m}{B(a,b)(a+m)m!}$ and $(f)_k = f(f+1)\dots(f+k-1)$.

Combining Equations (2.1) and (2.11), the pdf of the i th order statistic of the BPP distribution can be easily obtained by

$$f_{i:n}(x) = \sum_{m=0}^{\infty} k_{i,n,m} f_{BPP}(x, a + m, b, \gamma, \theta, \mu) \tag{2.12}$$

where

$$k_{i,n,m} = \sum_{j=0}^{n-i} \frac{\Gamma(b)\Gamma(n-i+1)\Gamma(a+m)(-1)^j}{j!\Gamma(n-i-j+1)\Gamma(a+b+m)B(i,n-i+1)} c_{i+j-1,m} \tag{2.13}$$

Combining Equations (2.9) and (2.12), the mgf of the i th order statistic of the BPP distribution, say $M_{i:n}(t)$, can be easily obtained by

$$M_{i:n}(t) = \frac{1}{e^\theta - 1} \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} \sum_{r=0}^{\infty} \frac{w_{i,a+m,b,\theta} (t\mu(\theta_j)^{\frac{1}{\gamma}})^r \varphi(\frac{\gamma-r}{\gamma}, \theta_j) k_{i,n,m}}{r!}$$

Using the Equations (2.10) and (2.12) we immediately obtain the r th generalized moment of $X_{i:n}$ as

$$E(X_{1:n}^r) = \frac{1}{e^\theta - 1} \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} \frac{w_{i,a+m,b,\theta} (t\mu(\theta_j)^{\frac{1}{\gamma}})^r \varphi(\frac{\gamma-r}{\gamma}, \theta_j) k_{i,n,m}}{r!}.$$

3 Estimation and inference

In this section, we discuss the estimation of the parameters of the BPP distribution. Let X_1, X_2, \dots, X_n be a random sample with observed values x_1, x_2, \dots, x_n from BPP distribution with parameters a, b, γ, θ and μ . Let $\Theta = (a, b, \gamma, \theta, \mu)^T$ be the parameter vector. The total log-likelihood function is given by

$$\begin{aligned} l_n &\equiv l_n(x; \Theta) = n \log \gamma + n \log \theta + n \log \mu + \theta \sum_{i=1}^n \left(\frac{\mu}{x_i}\right)^\gamma - (\gamma + 1) \sum_{i=1}^n \log x_i \\ &+ (a - 1) \sum_{i=1}^n \log (e^\theta - e^{\theta(\frac{\mu}{x_i})^\gamma}) + (b - 1) \sum_{i=1}^n \log (e^{\theta(\frac{\mu}{x_i})^\gamma} - 1) \\ &- n \log(e^\theta - 1) + n \log \Gamma(a + b) - n \log \Gamma(a) - n \log \Gamma(b) \end{aligned}$$

The associated score function is given by $U_n(\Theta) = (\frac{\partial l_n}{\partial a}, \frac{\partial l_n}{\partial b}, \frac{\partial l_n}{\partial \gamma}, \frac{\partial l_n}{\partial \theta}, \frac{\partial l_n}{\partial \mu})^T$, where

$$\begin{aligned} \frac{\partial l_n}{\partial \gamma} &= \frac{n}{\gamma} + \theta \sum_{i=1}^n \left(\frac{\mu}{x_i}\right)^\gamma \ln\left(\frac{\mu}{x_i}\right) - \sum_{i=1}^n \log x_i - (a - 1) \sum_{i=1}^n \frac{\theta(\frac{\mu}{x_i})^\gamma \ln(\frac{\mu}{x_i})}{e^\theta - e^{\theta(\frac{\mu}{x_i})^\gamma}} \\ &+ (b - 1) \sum_{i=1}^n \frac{\theta(\frac{\mu}{x_i})^\gamma \ln(\frac{\mu}{x_i})}{e^{\theta(\frac{\mu}{x_i})^\gamma} - 1} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial l_n}{\partial \theta} &= \frac{n}{\theta} + \sum_{i=1}^n \left(\frac{\mu}{x_i}\right)^\gamma + (a - 1) \sum_{i=1}^n \frac{e^\theta - (\frac{\mu}{x_i})^\gamma e^{\theta(\frac{\mu}{x_i})^\gamma}}{e^\theta - e^{\theta(\frac{\mu}{x_i})^\gamma}} \\ &- \frac{ne^\theta}{e^\theta - 1} + (b - 1) \sum_{i=1}^n \frac{(\frac{\mu}{x_i})^\gamma e^{\theta(\frac{\mu}{x_i})^\gamma}}{e^{\theta(\frac{\mu}{x_i})^\gamma} - 1} \end{aligned}$$

and

$$\frac{\partial l_n}{\partial a} = \sum_{i=1}^n \log (e^\theta - e^{\theta(\frac{\mu}{x_i})^\gamma}) + n\Psi(a + b) - n\Psi(a)$$

and

$$\frac{\partial l_n}{\partial b} = \sum_{i=1}^n \log (e^{\theta(\frac{\mu}{x_i})^\gamma} - 1) + n\Psi(a + b) - n\Psi(b)$$

and

$$\begin{aligned} \frac{\partial l_n}{\partial b} &= \theta \sum_{i=1}^n \frac{\gamma \theta}{x_i} \left(\frac{\mu}{x_i}\right)^{\gamma-1} - (a-1) \sum_{i=1}^n \frac{\gamma \theta}{x_i} \left(\frac{\mu}{x_i}\right)^{\gamma-1} \frac{e^{\theta(\frac{\mu}{x_i})^\gamma}}{e^\theta - e^{\theta(\frac{\mu}{x_i})^\gamma}} \\ &+ (b-1) \sum_{i=1}^n \frac{\gamma \theta}{x_i} \left(\frac{\mu}{x_i}\right)^{\gamma-1} \frac{e^{\theta(\frac{\mu}{x_i})^\gamma}}{e^{\theta(\frac{\mu}{x_i})^\gamma} - 1} \end{aligned}$$

As $\frac{\partial \log \Gamma(x)}{\partial x} = \Psi(x)$; that the Ψ is called digamma function. The MLE of Θ , say $\hat{\Theta}$, is obtained by solving the nonlinear system $U_n(\Theta) = 0$.

4 Applications of the BPP distribution

To show the superiority of the BPP distribution, we compare the results of fitting this distribution to some of their sub-models such as BP, BGP and PP distributions, using a real data set. The data set consists of 23 observations of ball bearing data from Caroni (2002). Table 1 lists the MLEs of the parameters, the values of AIC (Akaike Information Criterion), BIC (Bayesian information criterion) and $-2\log(L)$ for the first data. These values show that the BPP distribution provide a better fit than the BP, BGP and PP for fitting the first data. We apply the AIC and BIC statistics, in order to verify which distribution fits better to this data. In general, the smaller the values of AIC and BIC, the better the fit to the data. According to these statistics in Table 1, the BPP distribution fit the first data set better than the others. Using the likelihood ratio (LR) test, we test the null hypothesis $H_0 : PP$ versus the alternative hypothesis $H_1 : BPP$, or equivalently, $H_0 : (a, b) = (1, 1)$ versus the alternative hypothesis $H_1 : (a, b) \neq (1, 1)$, The value of the LR test statistic and the corresponding p-value are 58 and $2.54e - 13$, respectively. Therefore, the null hypothesis (PPmodel) is rejected in favor of the alternative hypothesis (BPP model) for any significance. For test the null hypothesis $H_0 : BP$ versus the alternative hypothesis $H_1 : BPP$, or equivalently, $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, the value of the LR test statistic is 19.4 ($p\text{-value} < 0.0001$), which includes that the null hypothesis (BP model) is rejected in favor of the alternative hypothesis (BPP model) for any significance. We also test the null hypothesis $H_0 : BGP$ versus the the alternative hypothesis $H_1 : BPP$, or equivalently, $H_0 : (b, \theta) = (1, 0)$ versus $H_1 : (b, \theta) \neq (1, 0)$. the value of the LR test statistic is 23.6 ($p\text{-value} < 0.0001$), which includes that the null hypothesis (BGP model) is rejected in favor of the alternative hypothesis (BPP model) for any significance level.

Table 1: MLEs, AIC, BIC, $-2\log(L)$ for the ball bearing

Distribution	MLEs	AIC	BIC	$-2\log(L)$
BPP	$\hat{\gamma} = 4.652, \hat{\theta} = 2.302, \hat{\mu} = 17.87, \hat{a} = 0.8274, \hat{b} = 0.0918$	266.8	272.5	256.8
PP	$\hat{\gamma} = 0.2381, \hat{\theta} = 0.0002, \hat{\mu} = 0.3239$	320.8	324.2	314.8
BP	$\hat{\gamma} = 0.4277, \hat{\mu} = 17.87, \hat{a} = 0.2725, \hat{b} = 0.5516$	284.2	288.7	276.2
BGP	$\hat{\gamma} = 1.830, \hat{\mu} = 1.627, \hat{a} = 500.46$	286.4	289.8	280.4

Conclusion

A new five-parameter distribution called the BPP distribution is introduced. Several properties of the new distribution are obtained. Fitting the BPP model to a real data set indicates the flexibility of the proposed distribution in data modeling. It seems that the proposed model can be considered as a suitable candidate model in reliability analysis.

References

- Akinsete, A., Famoye, F. and Lee, C. (2008). The beta-Pareto distribution, *Statistics: A Journal of Theoretical and Applied Statistics*, **42**, 547-563.
- Barreto-Souza, W., de Morais, A.L. and Cordeiro, G.M. (2011). The Weibull-geometric distribution, *Journal of Statistical Computation and Simulation*, **81**, 645-657.
- Cordeiro, G.M., Silva, G.O. and Ortega, E.M.M. (2011). The beta-Weibull geometric distribution, *Statistics: A Journal of Theoretical and Applied Statistics*, **47(4)**, 817-834.
- Cooray, K. (2010). Generalized Gumbel distribution, *Journal of Applied Statistics*, **37**, 171-179.
- Caroni, C. (2002). The correct ball bearings data, *Lifetime Data Analysis*, **8**, 395-399.
- Eugene, N., Lee, C. and Famoye, F. (2002). Beta-Normal distribution and its applications, *Communication in Statistics-Theory and Methods*, **31(4)**, 497-512.
- Gupta, R. D. and Kundu, D. (2001). Exponentiated exponential family: An alternative to gamma and Weibull distributions, *Australian New Zealand Journal of Statistics*, **41**, 173-188.
- Martines, E. S. and Stedinger, J. R. (2001). Generalized maximum likelihood Pareto-Poisson estimators for partial duration series, *Water Resources Research*, **32**, 2551-2557.

- Nadarajah, S. and Kotz, S. (2006). The beta exponential distribution, *Reliability Engineering and System Safety*, **91**, 689-697.
- Nadarajah, S. (2005). Exponentiated Pareto distributions, *Statistics: A Journal of Theoretical and Applied Statistics*, **39**, 255-260.
- Nassar, M. M. and Nada, N. K. (2011). The Beta Generalized Pareto distribution, *Journal of Statistics: Advances in Theory and Applications*, **6**, 1-17.
- Silva, G.O., Ortega, E.M. and Cordeiro, G.M. (2010). The beta modified Weibull distribution, *Lifetime Data Analysis*, **16**, 409-430.



Comparing Recurrent and Vector Singular Spectrum Analysis with Artificial Neural Networks in Time series Analysis

M. Yarmohammadi¹

¹Department of Statistics, Payame Noor University, Tehran, Islamic Republic of Iran

Abstract: In this research we consider three nonparametric techniques known as Artificial Neural Networks (NN) and Recurrent and Vector Singular Spectrum Analysis (SSA) for forecasting real time series data. As nonparametric techniques are not bound by any assumptions, it is more likely to provide an accurate representation of the true scenario in comparison to parametric techniques.

Keywords Artificial Neural Networks , Singular Spectrum Analysis , forecasting , Exchange rate.

1 Introduction

In this research we consider Artificial Neural Networks (NN) and Singular Spectrum Analysis (SSA) for forecasting exchange rates in the United Kingdom, European Union and China. The basic SSA method consists of two complementary stages: decomposition and reconstruction; both stages include two separate steps. At the first stage we decompose the series and at the second stage we reconstruct the original series and use the reconstructed series for forecasting new data points [Gdyan dinaand all \(2001\)](#). Artificial neural networks are generally presented as systems of interconnected "neurons" which can compute values from inputs. Like other machine learning methods systems that learn from data neural networks have been used to solve a wide variety of tasks that are hard to solve using ordinary rule based programming, including computer vision and speech recognition [Marzban,and Akbarian \(2005\)](#). The structure of the paper is as follows. The next section describes the SSA technique and provides some general rules for Forecasting time series data. The technique of artificial neural networks is introduced briefly. The time series data and empirical results are presented in Sections 4.

2 Singular Spectrum Analysis (SSA)

SSA is essentially a model-free technique; it is more an exploratory, model building tool than a confirmatory procedure. It aims at a decomposition of the original series into a sum of a small number of interpretable components such as a slowly varying trend, oscillatory components and a structureless noise. The SSA technique consist of two complementary stages: decomposition and reconstruction and

both of which include two separate steps [Gdyan dinaand all \(2001\)](#). In what follows, we give details of these steps.

Step 1 Embedding:

Consider the real-valued non-zero time series $Y_T = (y_1, \dots, y_T)$ of sufficient length T . Let $[K = T - L + 1]$ where L ($L \leq T/2$) is some integer called the window length. Define the trajectory matrix \mathbf{X} as follows:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & \cdots & \vdots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{pmatrix}$$

Note that \mathbf{X} is a Hankel matrix, which means that all the elements along the off diagonal are equal. We then consider \mathbf{X} as multivariate data with L characteristics and $[K = T - L + 1]$ observations. The columns $X_j = (y_j, \dots, y_{L+j-1})^T$ of \mathbf{X} , considered as vectors, lie in an L -dimensional space Re^L .

Step2: Singular Value Decomposition

Define the matrix $\mathbf{X}\mathbf{X}^T$. Singular value decomposition (SVD) of $\mathbf{X}\mathbf{X}^T$ provides us with the collections of L eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$ and the corresponding eigenvectors U_1, U_2, \dots, U_L , where U_i is the normalized eigenvector corresponding to the eigenvalue λ_i , ($i = 1, \dots, L$). Then SVD of the trajectory matrix \mathbf{X} can be written as:

$$\mathbf{X} = \mathbf{E}_1 + \dots + \mathbf{E}_L$$

Where, $\mathbf{E}_i = \sqrt{\lambda_i} U_i V_i^T$, $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$ ($i = 1, \dots, L$)

Step3: Grouping

Partition the set of indices ($i = 1, \dots, L$) into m disjoint subsets $\{I_1, \dots, I_m\}$. This

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}$$

Where, $\mathbf{X}_{I_j} = \sum_{\ell \in I_j} \mathbf{E}_\ell$, $j = 1, \dots, m$

Step4: Hankelization

The last step in SSA, transforms each matrix of the grouped decomposition (step3) into a new series of length T . This will be done with diagonal averaging as follows. If Z_{ij} stands for an element of a matrix Z , then the k -th term of the resulting series is obtained by averaging Z_{ij} over all i, j such that $i+j = k+1$. The result of the Hankelization of a matrix Z is the Hankel matrix \mathbf{HZ} , which is the trajectory matrix corresponding to the time series obtained as a result of the diagonal averaging. The operator

H acts on an arbitrary LK matrix $Z=Z_{ij}$ with $L \leq K$ in the following way: for $i+j =s$ and $N=L+K-1$ the element \tilde{z}_{ij} of the matrix HZ is:

$$\tilde{z}_{ij} = \begin{cases} \frac{1}{s-1} \sum_{l=1}^{s-1} z_{l,s-l} & , 2 \leq s \leq L-1; \\ \frac{1}{L} \sum_{l=1}^L z_{l,s-l} & , L \leq s \leq K+1; \\ \frac{1}{N-s+1} \sum_{l=s-K}^L z_{l,s-l} & , K+2 \leq s \leq N+1; \end{cases}$$

Another way of obtaining hankelized form of the matrix HZ is as bellows:

$$\tilde{z}_{ij} = \frac{1}{s_2 - s_1 + 1} \sum_{l=s_1}^{s_2} z_{l,s-l}$$

Where, $s_1 = \max \{1, s - N + L\}$, $s_2 = \min \{s, L\}$

By performing the diagonal averaging of all matrix components of \mathbf{X}_{I_j} in the expansion of \mathbf{X} above, we obtain another expansion:

$$\mathbf{X} = \tilde{\mathbf{X}}_{I_1} + \dots + \tilde{\mathbf{X}}_{I_m}$$

where $\tilde{\mathbf{X}}_{I_j}$ is the diagonalized version of the matrix \mathbf{X}_{I_j} . This is equivalent to the decomposition of the initial series $Y_T = (y_1, \dots, y_T)^T$ into a sum of m series; $y_t = \sum_{j=1}^m \tilde{y}_t^{(j)}$, where $\tilde{Y}_T^{(j)} = (\tilde{y}_1^{(j)}, \dots, \tilde{y}_T^{(j)})^T$

2.1 Forecasting by SSA

The basic requirement to make SSA forecasting is that the time series satisfies a linear recurrent formula (LRF). A time series $Y_T = (y_1, \dots, y_T)$ satisfies LRF of order d if:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_d y_{t-d}, t = d+1, \dots, T.$$

The time series governed by LRFs admit a natural recurrent continuation because each term of that time series is equal to a linear combination of several preceding terms. In what follows, we give a brief description of Recurrent SSA (RSSA) and Vector SSA (VSSA) algorithms. Further details can be found in [Gdyan dinaand all \(2001\)](#). Let us assume that U_j^∇ is the vector of the first $L-1$ components of the eigenvector U_j and π_j is the last component of U_j ($j = 1, \dots, r$). We can then define the coefficient vector Re as $\text{Re} = \frac{1}{1-v^2} \sum_{j=1}^r \pi_j U_j^\nabla$, where $v^2 = \sum_{j=1}^r \pi_j^2$

Recurrent SSA

Considering the above notation, the RSSA forecasts $(\hat{y}_{T+1}, \dots, \hat{y}_{T+M})$ can be obtained by

$$\hat{y}_i = \begin{cases} \tilde{y}_i, & i = 1, \dots, T \\ \text{Re}^T Z_i, & i = T + 1, \dots, T + M \end{cases},$$

where, $Z_i = [\hat{y}_{i-L+1}, \dots, \hat{y}_{i-1}]^T$ and $\tilde{y}_1, \dots, \tilde{y}_T$, are the values for the reconstructed time series and can be obtained from reconstruction stage, with $\mathbf{X}_{I_j} = \sum_{i=1}^r \mathbf{X}_i$

Vector SSA

Consider the following matrix $\mathbf{P} = \mathbf{U}^\nabla \mathbf{U}^{\nabla T} + (1 - v^2) \text{Re} \text{Re}^T$. Now, define the linear operator: $\mathcal{P}^{(v)} : \mathcal{L}_r \mapsto \text{Re}^L$, where $\mathcal{L}_r = \text{span}\{U_1, \dots, U_r\}$ and

$$\mathcal{P}^{(v)} Y = \begin{pmatrix} \mathbf{P} Y_\Delta \\ \text{Re}^T Y_\Delta \end{pmatrix}, Y \in \mathcal{L}_r.$$

Suppose the vector Z_i is defined as follows

$$Z_j = \begin{cases} \tilde{X}_j & \text{for } j = 1, \dots, K \\ \mathcal{P}^{(v)} Z_{j-1} & \text{for } j = K + 1, \dots, K + M + L - 1 \end{cases},$$

where \tilde{X}_j are the j^{th} reconstructed columns of the trajectory matrix of the time series after grouping and discarding noise components. Now, by constructing the matrix $Z = [Z_1, \dots, Z_{K+M+L-1}]$ and performing diagonal averaging, we obtain a new time series $\hat{y}_1, \dots, \hat{y}_{T+M+L-1}$, where $\hat{y}_{T+1}, \dots, \hat{y}_{T+M}$ form the terms of the VSSA forecast. For more details see [Gdyan dinaand all \(2001\)](#).

3 Artificial Neural Networks

Artificial neural network research stagnated after the publication of machine learning research by [Minsky and Papert \(1969\)](#). They discovered two key issues with the computational machines that processed neural networks. The second significant issue was that computers were not sophisticated enough to effectively handle the long run time required by large neural networks. Neural network research slowed until computers achieved greater processing power. Also key later advances was the back propagation algorithm which effectively solved the exclusive-or problem [Werbos \(1975\)](#). In supervised learning Which is consider here, we are given a set of example pairs $(x, y), x \in X, y \in Y$ and the aim is to find a function $f : X \rightarrow Y$ in the allowed class of functions that matches the examples. A commonly used cost is the mean-squared error, which tries to minimize the average squared error between the network's output, $f(x)$, and the target value y over all the example pairs [Priddy and Keller \(2005\)](#).

4 Application of SSA and ANN for Exchange rate data

Exchange rate between two currencies is regarded as the value of one country's currency in terms of another currency. The data which are used in this research relates to daily exchange rates in the United Kingdom, European Union and China between the dates 03/01/2012 and 01/03/2013. In order to train the forecasting models, approximately 2/3 rd of the observations (i.e. 192 observations) were used while approximately 1/3 rd of the observations (i.e. 100) were left aside for testing the forecasting accuracy of the models. The measures were adopted to compare and contrast between the forecasting models are and Direction of change(D).

4.1 Results

Table 1 reports the RMSEs for the out-of-sample forecasting results. All RMSEs have been multiplied by 103 in order to enable easy comparison between the forecasting models. Firstly, based on the RMSE, the results show that we cannot identify one model to be best for all three countries at all horizons. Instead, we see that for forecasting exchange rates in UK, the Basic VSSA model is able to outperform both NN and Basic RSSA at horizons of $h = 1$ and 3 steps ahead. For EU, we see that the NN model outperforms both Basic VSSA and Basic RSSA at forecasting the exchange rate at all horizons. The results for China show a completely different outlook as the NN models forecasting accuracy appears to have deteriorated to a great extent and both Basic VSSA and Basic RSSA models outperform the NN model. In China, for 1 day ahead forecasts of exchange rates the Basic VSSA model outperforms Basic RSSA marginally whilst for 3 day ahead forecasts, the Basic RSSA model is seen to outperform the Basic VSSA model marginally.

Table 1: Out of sample forecasting results for exchange rates .

1 Series	NN		Basic VSSA		Basic RSSA		$\frac{\text{BasicVSSA}}{\text{NN}}$		$\frac{\text{BasicVSSA}}{\text{BasicRSSA}}$	
	1	3	1	3	1	3	1	3	1	3
UK	7.62	12.8	6.80	11.5	7.10	11.7	0.89	0.89	0.96	0.98
EU	6.77	11.0	7.00	11.8	7.10	11.8	1.03	1.07	0.99	1.00
CHina	15.7	25.0	5.19	7.95	5.21	7.92	0.33**	0.32*	0.996	1.004

Note: (*) indicates statistical significance based on the DM test at $p=0.05$ and (**) $p=0.01$

Based on the RRMSE, we can see that only two outcomes are statistically significant. Accordingly, we can conclude with 95% confidence that the Basic VSSA model is 67% better than NN at forecasting the Chinese exchange rate at a horizon of one day ahead. Furthermore, we are able to conclude with 99% confidence that for 3 day ahead forecasts of Chinese exchange rates, the Basic VSSA model is

68% better than NN. Finally, we tested the ability of the forecasting models at predicting the actual direction of change in the exchange rate time series. The results are reported in Table 2. Based on the DC, for all three countries, the NN model provides the worst DC prediction in comparison to Basic VSSA and Basic RSSA models. For UK, the Basic VSSA model provides the best DC prediction at $h = 1$ and 3 days ahead. For EU, Basic VSSA has the most favorable DC prediction at one step ahead, but NN appears to have the best DC prediction ability at three steps ahead. Finally, for China we can see that Basic RSSA records the best DC prediction at $h = 1$ step ahead whilst Basic VSSA provides the best DC prediction at three steps ahead. Once again we test our results for statistical significance. Accordingly, the Table 2 shows that all conclusions on DC which were made previously can be attributed to chance occurrences. However, we are able to conclude with 95% confidence that for the Chinese exchange rate forecast, the NN model provides the worst DC prediction at 3 steps ahead with only 48% accuracy.

Table 2: Direction of change results for exchange rates .

	NN		Basic VSSA		Basic RSSA	
1	1	3	1	3	1	3
Series	1	3	1	3	1	3
UK	0.43	0.57	0.51	0.62	0.49	0.51
EU	0.48	0.69	0.52	0.51	0.46	0.51
CHina	0.52	0.48*	0.54	0.57	0.62*	0.54

Note *: indicates statistical significance based on the DM test at $p=0.05$

References

- Gdyan dina, N. Nekrutkin V. Zhigljavsky, A(2001). Analysis of time series :SSA and related techniques Chapman and Hall.
- Marzban, H., Akbarian, R., and Javaheri, B. (2005). Comparison of Structural econometric models, Time series and Artificial Neural Network for predicting exchange rate. *Journal of Economic Research*, **69**, 181–216.
- Minsky, M.; S. Papert (1969). An Introduction to Computational Geometry, *MIT Press*. ISBN , 0-262-63022-2.
- Werbos, P.J. (2003). New Tools for Prediction and Analysis in the Behavioral Sciences. *Beyond Regression*, .

Priddy, K. L. and Keller, P. E. (2005). Artificial Neural Networks: An Introduction, *JSPIE Press*, 165 pages.

Hassani H, Soofi, A. S, and Zhigljavsky, A A. (2010). Predicting daily exchange rate with singular spectrum analysis, *Nonlinear Analysis: Real World Applications*, **11**, 2023–2034.



Application of Statistical Tests of Anisotropy for Fractional Brownian Textures to Mammograms

Gholamhossein Yari, Soodeh Moallemian¹

School of Mathematics, Iran University of Science and Technology, Tehran, Iran

Abstract: In this paper, we apply a methodology for analysis of texture anisotropy. This method is based on some statistical tests which are able to detect anisotropy of a texture. The method is introduced and applied on simulated data. We show that on a database of 100 full-field mammograms, about 60 percent of textures can be considered as anisotropic with a high level of confidence. These empirical results suggest that the fractional Brownian fields are better than the usual Brownian fields for modeling the mammogram texture.

Keywords: Anisotropy, Anisotropic fractional Brownian field, Texture analysis, Hurst index, Mammography.

Mathematics Subject Classification (2010): 62M40 62P10 62H15.

1 Introduction

Medical imaging techniques have widely improved and revolutionized health care delivery all around the world. Melding the medical imaging and the power of Mathematics and Statistics offers highly personalized and targeted means of diagnosis or prognosis generation and is fostering higher quality and efficiency in health care.

One important aspect of a medical image is its texture which delivers information about the spatial arrangement of intensities in that image and is essential for processing the image see e.g. [Richard and Bierme \(2010\)](#). From a study to another, the definition of texture varies, depending mainly on analysis approaches. In such approaches, texture features are usually derived from the estimation of model parameters.

Elements of the textures we are dealing with, are satisfied in Fractional Brownian field. [Kolmogorov \(1940\)](#), [Mandelbrot and Van Ness \(1968\)](#) defined Fractional Brownian field (FBf) as the unique centered Gaussian field, with stationary increments, isotropic and self-similar of order $H \in (0, 1)$. In this work, we deal with rough textures which appear on irregular images. A Hölderian meaning is given to this irregularity. From this point of view, the textures are related to high frequency components of images,

¹Soodeh Moallemian : moallemian@mathdep.iust.ac.ir

and not associated to their possible trends and other low frequency ones. In Figure 1, you can see some images having our region of interest (ROI).

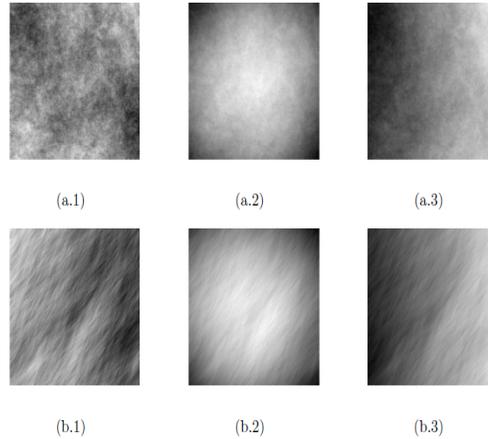


Figure 1: Isotropy and anisotropy of textures of images. (a.1)-(a.3) are isotropic and (b.1)-(b.3) are anisotropic.

The study of random field is a wide research field in Probability Theory and covers a lot of open problems related to its definition and anisotropy analysis, estimation of anisotropic model parameters and anisotropic field simulation see e.g. [Mandelbrot and Van Ness \(1968\)](#) etc.

This paper is organized as follows. In Section 2, we give an overall presentation of Gaussian random fields as well as a brief description of mammography. The tests that are used for testing the anisotropy of the image database will be introduced briefly in Section 3. In Section 4, a summary of numerical studies is brought and finally, we discuss briefly about the results of the tests in Section 5.

2 Materials and Methods

In this section some fundamental definitions and descriptions are given.

2.1 Definitions

Let (Ω, A, \mathbb{P}) be a probability space. A d -dimensional *random field* Z is a map from $\Omega \times \mathbb{R}^d$ into \mathbb{R} such that $Z(\cdot, X) := Z(X)$ is a random variable on Ω for all $x \in \mathbb{R}^d$. When $d = 1$, such a field is called *Random Process*. If any finite linear combination of a random field's associated random variables is a Gaussian variable, then the field is a *Gaussian* one. A centered Gaussian field is characterized by its covariance function:

$$(x, y) \longmapsto \text{cov}(Z(x), Z(y)).$$

The law of a centered Gaussian field Z with stationary increments is characterized by its variogram which is defined by:

$$\forall x \in \mathbb{R}^d, \quad \vartheta(x) = \mathbb{E}((Z(X) - Z(0))^2).$$

these fields have the variogram of the form:

$$\forall x \in \mathbb{R}^d, \quad \vartheta(x) = \int_{\mathbb{R}^d} |e^{ix \cdot \xi} - 1|^2 f(\xi) d\xi. \tag{2.1}$$

For $d = 2$ these fields are called Gausssian fields with *spectral density*. The fields we focus on are of this form with the following spectral density:

$$\forall \xi \in \mathbb{R}^2, \quad f(\xi) = |\xi|^{-2h(\arg(\xi)-2)}, \tag{2.2}$$

where h is a measurable π -periodic function with range $[H, M] \subset (0, 1)$ where $H = \text{essinf}_{[-\pi, \pi]} h$ and $M = \text{esssup}_{[-\pi, \pi]} h$. We refer to these fields as Extended Fractional Brownian Field (EFBF).

2.2 Mammography

Mammography is a type of breast imaging that helps the process of examination and detection of breast change by using a certain dose of x-ray system, before they are noticeable or visible. This technique provides two-dimensional information that allows the radiologist to make a diagnosis.

2.3 Parameter estimation

Theorem 2.1. *Let X be a GFSD (Gaussian Field spectral density) and $\beta \in (0, 1)$.*

a) *let $0 < \alpha \leq \gamma < 1$. If there exist $A, B_1, B_2 > 0$ and a positive measure subset E of the unit sphere S^{d-1} of \mathbb{R}^d such that for almost all $\xi \in \mathbb{R}^d$: (i) $|\xi| \geq A \Rightarrow |f(\xi)| \leq B_1 |\xi|^{-2\alpha-d}$ (ii) $|\xi| \geq A$, and $\frac{\xi}{|\xi|} \in E \Rightarrow |f(\xi)| \leq B_2 |\xi|^{-2\gamma-d}$*

then there exist $\delta > 0$ and $c_1, c_2 > 0$ such that for all $y \in \mathbb{R}^d$: (iii) $|y| \leq \delta \Rightarrow c_1 |y|^{2\gamma} \leq \vartheta(y) \leq c_2 |y|^{2\alpha}$

b) *if condition (iii) holds for any α, γ with $0 < \alpha \leq \beta \leq \gamma < 1$ then β is the critical holder exponent of X .*

Proof. See [Bonami and Estrade \(2003\)](#). □

Theorem 2.2. *Let $Y = Y(t); t \in \mathbb{R}$ be a Gaussian random process with spectral density. Let $\beta = n + s$, with $n \in \mathbb{N}$ and $s \in (0, 1)$.*

a) *let $0 < \alpha \leq \gamma < 1$. If there exist $A, B_1, B_2 > 0$ such that for almost all $\xi \in \mathbb{R}^d$: (i) $|\xi| \geq A \Rightarrow B_1 |\xi|^{-2\gamma-2n-2} \leq |f_Y(\xi)| \leq B_2 |\xi|^{-2\alpha-2n-1}$ (ii) The variogram ϑ_Y is of class C^{2n} in a neighborhood of 0. (iii) There exist $\delta > 0$ and $c_1, c_2 > 0$ such that for all $t \in \mathbb{R}$: $|t| \leq \delta \Rightarrow c_1 |t|^{2\gamma} \leq |\vartheta^{(2n)}(t) - \vartheta^{(2n)}(0)| \leq c_2 |t|^{2\alpha}$*

b) *if conditions (ii) and (iii) hold for any α, γ with $0 < \alpha \leq s \leq \gamma \leq 1$, then β is the critical Hölder exponent of the process Y .*

Proof. See [Bonami and Estrade \(2003\)](#). □

The critical Hölder exponent h of an EFBF (when $d = 2$) is equal to the minimal value H of $h \in [-\pi, \pi] : H = \text{essinf}_{[-\pi, \pi]} h$ and is called *Hurst index*.

The Hurst index $h(\theta)$ of an EFBF in a given direction θ can be deduced from the Hurst index of the projected field perpendicular to this direction [Bonami and Estrade \(2003\)](#). As a consequence, the problem of estimation of the directional Hurst index of an EFBF is reduced to the problem of estimating the Hurst indices of projected fields.

Many Hurst index estimators have been proposed in the literature. The maximum likelihood estimator and the related Whittle estimator are often used to analyze Fractional Brownian motion (fBm) with long range dependence ($H \in (\frac{1}{2}, 1)$). Other estimators are defined by filtering discrete observations of fBm sample path. This is the case for the wavelet-based estimators [Abry et al. \(1995\)](#) or generalized quadratic variations. In this work we will use the generalized quadratic variations proposed in [Richard and Bierme \(2010\)](#).

3 Anisotropy Tests

3.1 Definitions

Let X be an EFBF with a directional Hurst index $H = \text{essinf}_{[-\pi, \pi]} h$. The field is isotropic if $h \equiv H$ or at least it can be considered as isotropic if $H = \text{esssup}_{[-\pi, \pi]} h$. So, one could try to test:

$$\begin{aligned} H_0 & : \quad h \equiv H \quad (\text{isotropy}), \\ H_1 & : \quad \exists \theta_1 \neq \theta_2, \quad h(\theta_1) \neq h(\theta_2) \quad (\text{anisotropy}). \end{aligned}$$

Such a test requires estimation of h in all directions which implies the discretization of Radon transform in an arbitrary direction and consequently, interpolation is needed. Hence, this estimation can be biased. In order to avoid interpolations and have a reliable implementation of the test, [Richard and Bierme \(2010\)](#) restricted the test definitions to vertical and horizontal directions. Let $h_1 = h(\theta_1)$ and $h_2 = h(\theta_2)$ denote the Hurst indices in vertical and horizontal directions. We test

$$\begin{aligned} H_0 & : \quad h_1 \equiv h_2 \quad (\text{weak isotropy I}), \\ H_1 & : \quad h_1 \neq h_2 \quad (\text{anisotropy}). \end{aligned}$$

It is emphasized that the null hypothesis does not imply the isotropy of the field. All isotropic fields, however, satisfy this condition. Assuming h is continuous in a neighborhood of θ_1 and θ_2 , hypothesis

H_1 implies that $H = \text{essinf}_{[-\pi, \pi]} h \neq \text{esssup}_{[-\pi, \pi]} h$ and therefore the field is anisotropic. A statistic of this test is normally defined as

$$\hat{d} = |\hat{h}_1 - \hat{h}_2|, \tag{3.1}$$

where $\hat{h}_1 = \hat{h}(\theta_1)$ and $\hat{h}_2 = \hat{h}(\theta_2)$ are defined in Richard and Bierme (2010). We expect \hat{d} to be high when the field is anisotropic. Hence, the rejection interval associated to the test is of the form

$$R_1 = \{\hat{d} > c_1\}, \tag{3.2}$$

where c_1 is a positive constant. This first test is only able to detect the anisotropy of the fields which have different horizontal and vertical directional Hurst indices, and fails at the ones with the same Hurst indices in these two directions. In order to attenuate this drawback, a second test is proposed as follows

$$\begin{aligned} H_0 & : H = h_1 \equiv h_2 \quad (\text{weak isotropy II}), \\ H_1 & : H \neq h_1 \text{ or } H \neq h_2 \quad (\text{anisotropy}). \end{aligned}$$

In that case, the statistic is given by

$$\hat{\delta} = |\max(\hat{h}_1, \hat{h}_2), \hat{H}|, \tag{3.3}$$

where \hat{H} is an estimator of the minimal Hurst index. The rejection interval of the second test is often of the form

$$R_2 = \{\hat{\delta} > c_2\}. \tag{3.4}$$

It can be proved that these test statistics are convergent. Interested readers may see the proof in Richard and Bierme (2010).

Theorem 3.1. *Let X be an EFBBF with directional Hurst index h continuously differentiable in a neighborhood of $\theta_1 = (0, 1)$ and $\theta_2 = (1, 0)$. Then as $N \rightarrow \infty$, $\hat{h}_1 - \hat{h}_2 \rightarrow h_1 - h_2$ almost surely. Where $\hat{h}_1 = \hat{h}(\theta_1)$ and $\hat{h}_2 = \hat{h}(\theta_2)$ are estimators of $h_1 = h(\theta_1)$ and $h_2 = h(\theta_2)$. In addition, there exist a positive constant $\gamma^2(h_1, h_2)$ that only depends on h_1 and h_2 , such that, as $N \rightarrow \infty$, $\sqrt{N}(\hat{h}_1 - \hat{h}_2 - (h_1 - h_2)) \rightarrow N(0, \gamma^2(h_1, h_2))$ in distribution.*

This theorem certifies that under the null hypothesis H_0 of the first test, as $N \rightarrow \infty$ we have $\sqrt{N}\hat{d} \rightarrow |N(0, \gamma^2(h_1, h_2))|$ in distribution, while under the assumption H_1 , $\sqrt{N}\hat{d} \rightarrow +\infty$ almost surely.

4 Numerical study

We generated a dataset of synthetic EFBBF of size 256×256 from Digital Database for Screening Mammography (DDSM)¹¹. The rejection bounds have been calculated for each test as $c_1 = 1.96 * 0.15 \simeq$

¹¹<http://marathon.csee.usf.edu/Mammography/Database.html>

0.3 and $c_2 = 0.2$ as perposed in [Richard and Bierme \(2010\)](#). After setting the parameters, we applied anisotropy tests and reported the percentages of cases detected as isotropic in Table 1. On isotropic simulations, the tests yield few errors, but when the minimal Hurst index is higher, the results are somewhat better. Level of confidence of the test is about 95%. The mammogram anisotropy is further

Table 1: Bias of the test statistics \hat{d} and $\hat{\delta}$ obtained on Stein’s simulations of isotropic fractional Brownian fields. The value p is the percentage of simulations classified as anisotropic corresponding to the statistics.

h	$\hat{d} \pm \sigma$	p	$\hat{\delta} \pm \sigma$	p
0.1	0.0 ± 0.17	91	0.09 ± 0.07	91
0.2	0.01 ± 0.16	94	0.08 ± 0.06	93
0.3	0.0 ± 0.16	94	0.08 ± 0.06	96
0.4	0.0 ± 0.16	94	0.08 ± 0.06	93
0.5	0.01 ± 0.14	98	0.07 ± 0.05	97
0.6	0.01 ± 0.14	97	0.08 ± 0.05	97
0.7	0.0 ± 0.13	99	0.07 ± 0.05	99
T	0.0 ± 0.15	95	0.08 ± 0.06	95

confirmed by result shown in Figures 2. In Figure2(a), there are about 14 percent of anisotropy detection, while the second histogram shows that the second test is able to detect about 60% of anisotropic cases. Note that these results do not state that in our dataset, 60% of the textures are anisotropic and the rest are isotropic. They only mean that there is at least 60% of cases, which according to the EFBF model, can be considered as anisotropic with a confidence level of 95%.

5 Discussion and Conclusion

The density of a mammogram refers to the bright image aspect caused by the presence of fibro glandular tissues in the breast. At the end of 60’s, [J.N. Wolfe \(1967\)](#) put forth the idea that the breast cancer risk could be assessed from the observation of mammogram appearance and patterns. Later, some investigators started focusing on the relationship between breast density and breast cancer risk; see e.g. [Boyd et al. \(1982\)](#) and [Brisson et al. \(1982\)](#). They provided the first evidence that any increase in breast density is related to an increase in cancer risk. This evidences are confirmed by many studies. Measurement of Hurst-related parameters on mammograms have been reported in many papers. The present work show that EFBF model enables to extract some significant density features which are not captured by the usual fractional Brownian field model. We have constructed a new method of

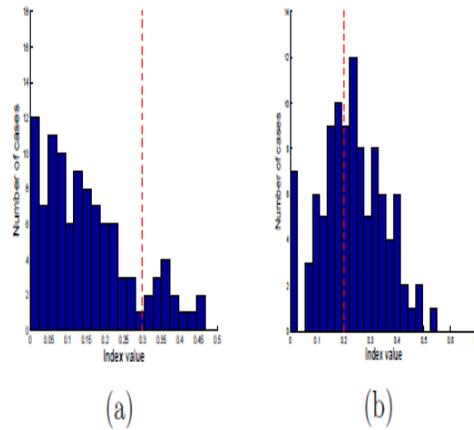


Figure 2: Histogram of estimators (a) $\hat{d} = |\hat{h}_1 - \hat{h}_2|$ and (b) $\hat{\delta} = |\max(\hat{h}_1, \hat{h}_2) - \min(\hat{h}_1, \hat{h}_2)|$. The dashed lines represent the rejection bounds of anisotropy test corresponding to each estimator.

extracting ROI (Region of Interest) using MATLAB programming. The statistical test and the new scripts were validated on simulated data.

References

- Abry, P., Goncalves, P. and Sellan, F. (1995). *Wavelet spectrum analysis and 1/f processes*, in: Lecture Notes in statistics, vol. 103, pp. 15-30. Springer, Berlin 1996.
- Bierme, H, Richard, F. (2008). *Estimation of anisotropic gaussian fields through radon transform.*, in: ESAIM: Probab. Stat., 12(1): 3050, 2008.
- Bonami, A, Estrade, A. (2003). *Anisotropic analysis of some Gaussian models.*, in: J. Fourier Anal. Appl., 9:215236, 2003.
- Boyd, N.-F., OSullivan, B., and Campbell, J.-E. (1982). *Mammographic signs as risk factors for breast cancer.*, in: Br. J. Cancer, 45:185193, 1982.
- Brisson, J, Merletti, F., Sadowsky, N.-L., et al. (1982). *Automated analysis of mammographic densities.*, in: Phys. Med. Biol., 41:909923, 1996.
- Demagnet, L., Ying, L. (2007). *Wave atoms and sparsity of oscillatory patterns.*, in: Appl. JHarmon. Anal., 23:368-387, 2007.

Kestener, P., Lina, J.-M., Saint-Jean, et al (2001). *Wienersche Spiralen und einige andere interessante Kurven in Hilbertsche Raum*, in: C. R. (Dokl.) Acad. Sci. URSS, 26:115-118, 1940.

Mandelbrot, B.-B., Van Ness, j. (1968). *Fractional Brownian motion, fractionnal noises and applications.*, in: SIAM Rev., 10:422-437, 1968.

Richard, F.-J.-P., Bierme, H. (2010). *Statistical tests of anisotropy for fractional brownian textures. application to full-field digital mammography.*, in: J. Math. Imaging Vis., 36(3):227-240, 2010.

J.N. Wolfe. *A study of breast parenchyma by mammography in the normal woman and those with benign and malignant disease.* Radiology, 89:201205, 1967.



Exponential Pareto II Distribution

Gholamhossein Yari¹, Zahra Tondpour²

School of Mathematics, Iran University of Science and Technology.

Abstract: In this article, we introduce a distribution, namely, exponential Pareto II which is generalization of exponential Pareto distribution. Various properties of the exponential ParetoII distribution such that moments, limiting behavior, mode, hazard function, shannon entropy and entropy for i^{th} order statistics of this distribution corresponding to the independent random samples size n are obtained. Finally a real data set is given to illustrate the use of this distribution.

Keywords Exponential distribution, Pareto distribution, Moment, Entropy, Order statistics.

Mathematics Subject Classification (2010): 62E15.

1 Introduction

Abed Al-kadim and Boshi (2013) proposed a new kind of distribution as the form

$F(x) = \int_a^{\frac{1}{1-F^\neq(x)}} f^\delta(t)dt$, in which $F^\neq(x)$ is cumulative distribution function (cdf) of any random variable X that follows any distribution like Normal, Exponential or Gamma,...atc and $f^\delta(t)$ is probability density function (pdf) of any random variable T . They studied this kind of distribution where $F^\neq(x)$ is the cdf of Pareto I distribution and $f^\delta(t)$ is the pdf of exponential distribution.

In this article, we study this kind of distribution where $F^\neq(x)$ is the cdf of Pareto II distribution and $f^\delta(t)$ is the pdf of exponential distribution.

2 The Probability Density Function of EPD II

Let $F^\neq(x)$ be the cdf of Pareto II distribution, $F^\neq(x) = 1 - (1 + \frac{x-\sigma}{p})^{-\theta}$, and $f^\delta(t)$ be the pdf of exponential distribution, $f^\delta(t) = \lambda e^{-\lambda t}$. The cdf of exponential Pareto II distribution (EPD II) is given by

$$F(x) = \int_0^{\frac{1}{1-F^\neq(x)}} f^\delta(t)dt = \int_0^{\frac{1}{1-(1+\frac{x-\sigma}{p})^{-\theta}}} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda(1+\frac{x-\sigma}{p})^\theta}, \quad (2.1)$$

where $\lambda, p, \theta > 0, \sigma \in \mathbb{R}$. Hence, the pdf of EPD II is given by:

$$f(x) = \frac{\lambda\theta}{p} \left(1 + \frac{x-\sigma}{p}\right)^{\theta-1} e^{-\lambda(1+\frac{x-\sigma}{p})^\theta} I_{(\sigma-p, \infty)}(x). \quad (2.2)$$

²Zahra tondpour: ztondpoor@mathdep.iust.ac.ir

The EPD II distribution reduces to the exponential distribution where $\theta = \sigma = p = 1$.

By setting $\sigma = p$ in relations (2.1) and (2.2), we obtain, $F(x) = 1 - e^{-\lambda(\frac{x}{p})^\theta}$ and

$f(x) = \frac{\lambda\theta}{p}(\frac{x}{p})^{\theta-1}e^{-\lambda(\frac{x}{p})^\theta}I_{(0,\infty)}(x)$ as the cdf and the pdf of the exponential Pareto distribution which is introduced by Abed Al-kadim and Boshi (2013).

3 Properties of the Exponential pareto II distribution

The hazard function associated with the EPD II is

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\lambda\theta}{p} \left(1 + \frac{x - \sigma}{p}\right)^{\theta-1} \quad x > \sigma - p$$

The limiting behaviors of exponential Pareto II density function and hazard function are given in the following lemma.

Lemma 3.1. *The limit of the exponential Pareto II density function as $x \rightarrow \infty$ is 0 and the limit of the exponential Pareto II hazard function as $x \rightarrow \infty$ and the limit as $x \rightarrow \sigma - p^+$ are given by*

$$\lim_{x \rightarrow \infty} h(x) = \begin{cases} \infty & \theta > 1 \\ \frac{\lambda}{p} & \theta = 1 \\ 0 & \theta < 1 \end{cases} \quad \text{and} \quad \lim_{x \rightarrow \sigma - p^+} f(x) = \lim_{x \rightarrow \sigma - p^+} h(x) = \begin{cases} 0 & \theta > 1 \\ \frac{\lambda}{p} & \theta = 1 \\ \infty & \theta < 1 \end{cases}$$

Proof. The limit of the exponential Pareto II density function as $x \rightarrow \infty$ can be shown by using L'Hopital's rule. Other results are easy to see. \square

The following theorem shows that the exponential Pareto II is unimodal if $\theta \geq 1$ and doesn't have any defined mode if $\theta < 1$.

Theorem 3.2. *The exponential Pareto II distribution has a unique mode given by*

$$\text{mode} = \left(\frac{\theta - 1}{\theta\lambda}\right)^{\frac{1}{\theta}} p - p + \sigma$$

if $\theta \geq 1$ and doesn't have any defined mode if $\theta < 1$.

Proof. The derivative with respect to x of $\ln f(x)$ where $f(x)$ is pdf of EPD II is given by

$$\frac{\partial}{\partial x} \ln f(x) = \frac{\theta - 1}{x + p - \sigma} - \frac{\theta\lambda}{p} \left(1 + \frac{x - \sigma}{p}\right)^{\theta-1}.$$

Thus the critical point of $\ln f(x)$ is $x = \left(\frac{\theta-1}{\theta\lambda}\right)^{\frac{1}{\theta}} p - p + \sigma$. Now, for $\theta < 1$ since $\lim_{x \rightarrow \sigma - p^+} f(x) = \infty$ thus EPD II doesn't have any defined mode if $\theta < 1$.

Now, if $\theta \geq 1$, it is easy to see that $\frac{\partial^2}{\partial x^2} \ln f\left(\left(\frac{\theta-1}{\theta\lambda}\right)^{\frac{1}{\theta}} p - p + \sigma\right) < 0$. Hence $x = \left(\frac{\theta-1}{\theta\lambda}\right)^{\frac{1}{\theta}} p - p + \sigma$ is maximum point of $f(x)$ and since $\frac{\partial^2}{\partial x^2} \ln f(x)$ hasn't no root thus the EPD II has unique mode. \square

In figure 1, various graphs of the exponential Pareto II pdf are provided for different parameter values.

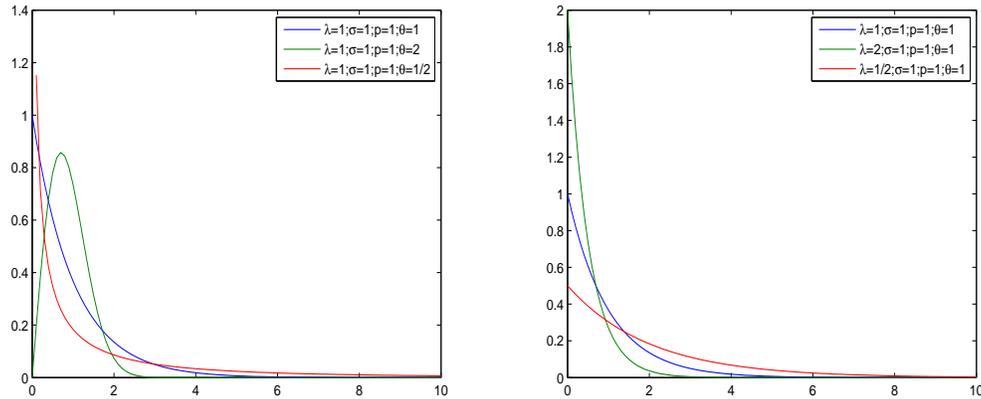


Figure 1: graphs of the exponential Pareto II pdf for various choices of λ, σ, p and θ

4 Moment

Theorem 4.1. *The r^{th} moment about the mean of exponential Pareto II distribution is given by*

$$E((X - \mu)^r) = \sum_{i=0}^r \binom{r}{i} \left(\frac{p}{\lambda^{\frac{1}{\theta}}}\right)^i (\sigma - \mu - p)^{r-i} \Gamma\left(\frac{i}{\theta} + 1\right) \quad r = 1, 2, 3, \dots$$

Proof.

$$E((X - \mu)^r) = \int_{\sigma-p}^{\infty} (x - \mu)^r \frac{\lambda\theta}{p} \left(1 + \frac{x - \sigma}{p}\right)^{\theta-1} e^{-\lambda\left(1 + \frac{x-\sigma}{p}\right)^\theta} dx,$$

by using the change of variable, $t = \lambda\left(1 + \frac{x-\sigma}{p}\right)^\theta, 0 < t < \infty$ we obtain,

$$\begin{aligned} E((X - \mu)^r) &= \int_0^\infty \left(\sigma - \mu - p + p\left(\frac{t}{\lambda}\right)^{\frac{1}{\theta}}\right)^r e^{-t} dt = \sum_{i=0}^r \binom{r}{i} (\sigma - \mu - p)^{r-i} \left(\frac{p}{\lambda^{\frac{1}{\theta}}}\right)^i \int_0^\infty t^{\frac{i}{\theta}} e^{-t} dt \\ &= \sum_{i=0}^r \binom{r}{i} (\sigma - \mu - p)^{r-i} \left(\frac{p}{\lambda^{\frac{1}{\theta}}}\right)^i \Gamma\left(\frac{i}{\theta} + 1\right). \end{aligned}$$

□

5 Entropy

The entropy of a random variable X is a measure of variation of uncertainty (Renyi, 1961). Shannon's entropy (Shannon, 1948) for a random variable X with pdf $f(x)$ is defined as $E(-\log(f(x)))$.

Theorem 5.1. *The Shannon's entropy of EPD II is given by*

$$H(X) = -\ln \frac{\lambda\theta}{p} + \frac{1-\theta}{\theta}(-\ln \lambda + \Gamma'(1)) + 1.$$

Proof.

$$H(X) = -\int_{\sigma-p}^{\infty} f(x) \ln f(x) dx = -\ln \frac{\lambda\theta}{p} - (\theta-1)E\left(\ln\left(1 + \frac{X-\sigma}{p}\right)\right) + \lambda E\left(\left(1 + \frac{X-\sigma}{p}\right)^\theta\right). \quad (5.1)$$

We need to find the expressions $E\left(\ln\left(1 + \frac{X-\sigma}{p}\right)\right)$ and $E\left(\left(1 + \frac{X-\sigma}{p}\right)^\theta\right)$. First we calculate the expectation of $\left(1 + \frac{X-\sigma}{p}\right)^\theta$.

$$E\left(\left(1 + \frac{X-\sigma}{p}\right)^\theta\right) = \int_{\sigma-p}^{\infty} \frac{\lambda\theta}{p} \left(1 + \frac{x-\sigma}{p}\right)^{2\theta-1} e^{-\lambda\left(1 + \frac{x-\sigma}{p}\right)^\theta} dx,$$

by using the change of variable, $t = \lambda\left(1 + \frac{x-\sigma}{p}\right)^\theta$, $0 < t < \infty$, we obtain,

$$E\left(\left(1 + \frac{X-\sigma}{p}\right)^\theta\right) = \int_0^\infty \frac{t}{\lambda} e^{-t} dt = \frac{1}{\lambda} \Gamma(2) = \frac{1}{\lambda}. \quad (5.2)$$

Now, we calculate $E\left(\left(1 + \frac{X-\sigma}{p}\right)^r\right)$

$$E\left(\left(1 + \frac{X-\sigma}{p}\right)^r\right) = \int_{\sigma-p}^{\infty} \frac{\lambda\theta}{p} \left(1 + \frac{x-\sigma}{p}\right)^{r+\theta-1} e^{-\lambda\left(1 + \frac{x-\sigma}{p}\right)^\theta} dx,$$

by using the change of variable, $t = \lambda\left(1 + \frac{x-\sigma}{p}\right)^\theta$, $0 < t < \infty$, we obtain

$$E\left(\left(1 + \frac{X-\sigma}{p}\right)^r\right) = \int_0^\infty \left(\frac{t}{\lambda}\right)^{\frac{r}{\theta}} e^{-t} dt = \lambda^{-\frac{r}{\theta}} \Gamma\left(\frac{r}{\theta} + 1\right). \quad (5.3)$$

Differentiating both sides of (5.3) with respect to r leads to,

$$E\left(\left(1 + \frac{X-\sigma}{p}\right)^r \ln\left(1 + \frac{X-\sigma}{p}\right)\right) = -\frac{1}{\theta} \lambda^{-\frac{r}{\theta}} \ln \lambda \Gamma\left(\frac{r}{\theta} + 1\right) + \frac{1}{\theta} \lambda^{-\frac{r}{\theta}} \Gamma'\left(\frac{r}{\theta} + 1\right). \quad (5.4)$$

Using relation (5.4) and $r = 0$, we obtain

$$E\left(\ln\left(1 + \frac{X-\sigma}{p}\right)\right) = -\frac{1}{\theta} \ln \lambda + \frac{1}{\theta} \Gamma'(1). \quad (5.5)$$

Substitute (5.2) and (5.5) in relation (5.1) implies that $H(X) = -\ln \frac{\lambda\theta}{p} + \frac{1-\theta}{\theta}(-\ln \lambda + \Gamma'(1)) + 1$. \square

6 Entropy for order statistics

Let X_1, X_2, \dots, X_n be a random sample of the EPD II and let $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$ denotes the corresponding order statistics; then

$$g_{i:n}(y) = n \binom{n-1}{i-1} (F_{i:n}(y))^{i-1} (1-F_{i:n}(y))^{n-i} f_X(y) = \frac{n\lambda\theta}{p} \binom{n-1}{i-1} (1-e^{-\lambda(1+\frac{y-\sigma}{p})^\theta})^{i-1} (1+\frac{y-\sigma}{p})^{\theta-1} e^{-\lambda(n-i+1)(1+\frac{y-\sigma}{p})^\theta} \quad y > \sigma - p$$

Theorem 6.1. *The Shannon's entropy for i^{th} order statistics of EPD II corresponding to the independent random samples size n is given by*

$$H_{i:n}(Y) = -\ln\left(\frac{n\lambda\theta}{p} \binom{n-1}{i-1}\right) - (i-1)\left(\frac{\Gamma'(i)}{(i-1)!} - \frac{\Gamma'(n+1)}{n!}\right) - \frac{n(\theta-1)}{\theta} \binom{n-1}{i-1} \\ \sum_{k=0}^{i-1} \binom{i-1}{k} (-1)^k \left[\frac{-\ln\lambda}{n-i+k+1} + \frac{\Gamma'(1) - (n-i+k+1)\ln(n-i+k+1)}{(n-i+k+1)^2} \right] + (n-i+1) \sum_{k=n-i+1}^n \frac{1}{k}. \tag{6.1}$$

Proof. The entropy expression of the $g_{i:n}(y)$ is:

$$H_{i:n}(Y) = -\int_{\sigma-p}^{\infty} g_{i:n}(y) \ln g_{i:n}(y) dy = -\ln\left(\frac{n\lambda\theta}{p} \binom{n-1}{i-1}\right) - (i-1)E(\ln(1 - e^{-\lambda(1+\frac{Y-\sigma}{p})^\theta})) - \\ (\theta-1)E(\ln(1 + \frac{Y-\sigma}{p})) + \lambda(n-i+1)E\left(\left(1 + \frac{Y-\sigma}{p}\right)^\theta\right). \tag{6.2}$$

We need to calculate the expression $E(\ln(1 + \frac{Y-\sigma}{p}))$, $E((1 + \frac{Y-\sigma}{p})^\theta)$ and $E(\ln(1 - e^{-\lambda(1+\frac{Y-\sigma}{p})^\theta}))$. First we calculate the expectation of $(1 + \frac{Y-\sigma}{p})^\theta$.

$$E\left(\left(1 + \frac{Y-\sigma}{p}\right)^\theta\right) = \frac{n\lambda\theta}{p} \binom{n-1}{i-1} \int_{\sigma-p}^{\infty} (1 - e^{-\lambda(1+\frac{y-\sigma}{p})^\theta})^{i-1} \left(1 + \frac{y-\sigma}{p}\right)^{2\theta-1} e^{-\lambda(n-i+1)(1+\frac{y-\sigma}{p})^\theta} dy,$$

by using the change of variable, $t = 1 - e^{-\lambda(1+\frac{y-\sigma}{p})^\theta}$, $0 < t < 1$ and then based on choosing $t = 1 - x$, $0 < x < 1$, we have

$$E\left(\left(1 + \frac{Y-\sigma}{p}\right)^\theta\right) = -\frac{n}{\lambda} \binom{n-1}{i-1} \int_0^1 (1-x)^{i-1} x^{n-i} \ln x dx = -\frac{n}{\lambda} \binom{n-1}{i-1} \frac{\Gamma(n-i+1)\Gamma(i)}{\Gamma(n+1)} \\ (\psi(n-i+1) - \psi(n+1)) = -\frac{1}{\lambda} (\psi(n-i+1) - \psi(n+1)).$$

Now, since $\psi(n+1) = -\gamma + \sum_{k=1}^n \frac{1}{k}$ where γ is Euler constant ($\gamma \simeq 0.5772156649\dots$) thus

$$E\left(\left(1 + \frac{Y - \sigma}{p}\right)^\theta\right) = \frac{1}{\lambda} \sum_{k=n-i+1}^n \frac{1}{k}. \quad (6.3)$$

Now, we calculate $E\left(\left(1 + \frac{Y - \sigma}{p}\right)^r\right)$

$$E\left(\left(1 + \frac{Y - \sigma}{p}\right)^r\right) = \frac{n\lambda\theta}{p} \binom{n-1}{i-1} \int_{\sigma-p}^{\infty} (1 - e^{-\lambda(1 + \frac{y-\sigma}{p})^\theta})^{i-1} \left(1 + \frac{y - \sigma}{p}\right)^{r+\theta-1} e^{-\lambda(n-i+1)(1 + \frac{y-\sigma}{p})^\theta} dy,$$

by using the change of variable, $t = 1 - e^{-\lambda(1 + \frac{y-\sigma}{p})^\theta}$, $0 < t < 1$, we have

$$E\left(\left(1 + \frac{Y - \sigma}{p}\right)^r\right) = n \binom{n-1}{i-1} \int_0^1 t^{i-1} (1-t)^{n-i} \left(-\frac{\ln(1-t)}{\lambda}\right)^{\frac{r}{\theta}} dt$$

and based on choosing $t = 1 - x$, $0 < x < 1$, and then by using the change of variable, $x = e^{-z}$, $0 < z < \infty$, we have,

$$\begin{aligned} E\left(\left(1 + \frac{Y - \sigma}{p}\right)^r\right) &= n \binom{n-1}{i-1} \lambda^{-\frac{r}{\theta}} \sum_{k=0}^{i-1} \binom{i-1}{k} (-1)^k \int_0^{\infty} e^{-(n-i+k+1)z} z^{\frac{r}{\theta}} dz \\ &= n \binom{n-1}{i-1} \lambda^{-\frac{r}{\theta}} \sum_{k=0}^{i-1} \binom{i-1}{k} (-1)^k \frac{\Gamma(\frac{r}{\theta} + 1)}{(n-i+k+1)^{\frac{r}{\theta}+1}} \end{aligned} \quad (6.4)$$

Differentiating both sides of (6.4) with respect to r leads to,

$$\begin{aligned} E\left(\left(1 + \frac{Y - \sigma}{p}\right)^r \ln\left(1 + \frac{Y - \sigma}{p}\right)\right) &= \frac{n}{\theta} \binom{n-1}{i-1} \lambda^{-\frac{r}{\theta}} \sum_{k=0}^{i-1} \binom{i-1}{k} (-1)^k \left[\frac{-\ln \lambda \Gamma(\frac{r}{\theta} + 1)}{(n-i+k+1)^{\frac{r}{\theta}+1}} + \right. \\ &\quad \left. \frac{\Gamma'(\frac{r}{\theta} + 1) - \Gamma(\frac{r}{\theta} + 1)(n-i+k+1)^{\frac{r}{\theta}+1} \ln(n-i+k+1)}{(n-i+k+1)^{\frac{2r}{\theta}+2}} \right], \end{aligned} \quad (6.5)$$

using relation (6.5) and $r = 0$, we obtain

$$\begin{aligned} E\left(\ln\left(1 + \frac{Y - \sigma}{p}\right)\right) &= \frac{n}{\theta} \binom{n-1}{i-1} \sum_{k=0}^{i-1} \binom{i-1}{k} (-1)^k \left[\frac{-\ln \lambda}{n-i+k+1} + \right. \\ &\quad \left. \frac{\Gamma'(1) - (n-i+k+1) \ln(n-i+k+1)}{(n-i+k+1)^2} \right]. \end{aligned} \quad (6.6)$$

Now, we calculate $E\left(\left(1 - e^{-\lambda(1 + \frac{Y - \sigma}{p})^\theta}\right)^r\right)$.

$$E\left(\left(1 - e^{-\lambda(1 + \frac{Y - \sigma}{p})^\theta}\right)^r\right) = \frac{n\lambda\theta}{p} \binom{n-1}{i-1} \int_{\sigma-p}^{\infty} (1 - e^{-\lambda(1 + \frac{y-\sigma}{p})^\theta})^{r+i-1} \left(1 + \frac{y - \sigma}{p}\right)^{\theta-1} e^{-\lambda(n-i+1)(1 + \frac{y-\sigma}{p})^\theta} dy,$$

by using the change of variable, $t = 1 - e^{-\lambda(1+\frac{y-\sigma}{p})^\theta}$, $0 < t < 1$, we have

$$\begin{aligned} E((1 - e^{-\lambda(1+\frac{y-\sigma}{p})^\theta})^r) &= n \binom{n-1}{i-1} \int_0^1 t^{i+r-1} (1-t)^{n-i} dt \\ &= n \binom{n-1}{i-1} \frac{\Gamma(i+r)\Gamma(n-i+1)}{\Gamma(n+r+1)}. \end{aligned} \tag{6.7}$$

Differentiating both sides of (6.7) with respect to r leads to,

$$\begin{aligned} E((1 - e^{-\lambda(1+\frac{y-\sigma}{p})^\theta})^r \ln(1 - e^{-\lambda(1+\frac{y-\sigma}{p})^\theta})) &= n \binom{n-1}{i-1} \left[\frac{\Gamma'(i+r)\Gamma(n-i+1)}{\Gamma(n+r+1)} - \right. \\ &\quad \left. \frac{\Gamma(i+r)\Gamma(n-i+1)\Gamma'(n+r+1)}{\Gamma^2(n+r+1)} \right], \end{aligned} \tag{6.8}$$

using relation (6.8) and $r = 0$, we obtain

$$E(\ln(1 - e^{-\lambda(1+\frac{y-\sigma}{p})^\theta})) = \frac{\Gamma'(i)}{(i-1)!} - \frac{\Gamma'(n+1)}{n!}. \tag{6.9}$$

Substitute (6.3), (6.6) and (6.9) in relation (6.2) implies (6.1). □

7 Application of the EPD II

In this section, the EPD II is applied to model a real data set. The required numerical evaluations are implemented using matlab.

We consider an uncensored data set from Birnbaum and Saunders (1969) on the fatigue life of 6061-T6 aluminum coupons cut parallel with the direction of rolling and oscillated at 18 cycles per second. The data set consists of 101 observations with maximum stress per cycle 31000 psi. The MLEs, the values of K-S (Kolmogorov-Smirnov) statistics with its respective p-value and AIC (Akaike Information Criterion) are listed in table 1. From these values, we conclude that the EPD II model provides a better fit to this data set than the EP, GP, Weibull and Pareto models.

Conclusion

In this article, we introduced a distribution, namely, exponential Pareto II distribution which is generalization of exponential Pareto distribution and various properties of this distribution are obtained. We obtained the exact form of the entropy expression for this distribution and the entropy for i^{th} order statistics of this distribution corresponding to the independent random samples size n . An application to a real data set indicates that the fit of the EPD II model is superior to the fits of EP, GP, Weibull and Pareto.

Table 1: MLEs, K-S statistics, p-value, AIC for the fatigue life of 6061-T6 Alminum coupons data

dist	MLE	K-S	p-value	AIC
EP II	$\hat{\lambda} = 2.3242, \hat{p} = 103.1562, \hat{\sigma} = 163.8421, \hat{\theta} = 3.4717$	0.0744	0.6050	924.5173
Weibull	$\hat{s} = 143.1661, \hat{k} = 6.0733$	0.099	0.2578	928.6291
EP	$\hat{\lambda} = 5.2910 \times 10^{-8}, \hat{p} = 3.7320, \hat{\theta} = 4.5732$	0.1264	0.0724	947.0739
GP	$\hat{a} = 299.9977, \hat{k} = 1.4151$	0.3272	0.0000	1075.9
Pareto	$\hat{\alpha} = 1.579, \hat{\theta} = 70$	0.3972	0.0000	1100

Appendix A

The distributions of EP, GP, Weibull and Pareto used in tables 1 are given by:

EP (Exponential Pareto): $f(x) = \frac{\lambda\theta}{p} \left(\frac{x}{p}\right)^{\theta-1} e^{-\lambda\left(\frac{x}{p}\right)^\theta}, \quad x > 0$

GP (Generalized Pareto): $f(x) = \frac{1}{a} \left(1 - \frac{kx}{a}\right)^{\frac{1-k}{k}}, \quad \begin{cases} 0 \leq x < \infty & k \leq 0 \\ 0 \leq x \leq \frac{a}{k} & k > 0 \end{cases}$

Weibull: $f(x) = \frac{k}{s} \left(\frac{x}{s}\right)^{k-1} e^{-\left(\frac{x}{s}\right)^k}, \quad x \geq 0$

Pareto: $f(x) = \frac{\alpha}{\theta} \left(\frac{x}{\theta}\right)^{-\alpha-1}, \quad x > \theta$

References

- Abed Al-kadim, K. and Boshi, M. (2013). Exponential pareto distribution, *Mathematical Theory and Modeling*, Vol.3, No.5, 135-146.
- Birnbaum, Z. W., Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue, *Journal of Applied Probability*, 6(1969) 328-347.
- Renyi, A. (1961). On measures of entropy and information, *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1(1961) 547-561.
- Shannon, C. E. (1948). A mathematical theory of communication, *Bell Technical Journal*, 27(1948) 379-432.



Shrinkage Estimation of Scale Parameters for Two Exponential Distributions Based on Record Values

H. Zakerzadeh, M. Karimi¹

Department of Statistics, Yazd University, Yazd, Iran

Abstract: The exponential distribution is applied in a very wide variety of statistical procedures. Among the most prominent applications are those in the field of life testing and reliability theory. When there are two record samples available for estimating the scale parameter, a preliminary test is usually used to determine whether to pool the samples or use the individual sample. In this paper, the preliminary test estimator and shrinkage estimator are studied. The optimum level of significance for preliminary test estimation and the optimum values of shrinkage coefficient are obtained based on minimax regret criterion under weighted square error loss function.

Keywords Minimax regret criterion, Optimal significance level, Preliminary test estimation, Record values, Shrinkage estimator.

Mathematics Subject Classification (2010): 62C20.

1 Introduction

Record values are of interest and important in many real life applications involving data relating to meteorology, sport, economics and life testing. Sometimes, the experimenter has some knowledge about the parameter of interest, either from past experience, or from experience with similar situations. This prior information may be incorporated in the estimation process using a preliminary test estimator and improved the estimation process.

Let X_1, X_2, \dots be a sequence of independent and identically distributed random variables having the same distribution. An observation X_j will be called an upper record value if exceeds in value all of the preceding observations, i.e., if $X_j > X_i$, for every $i < j$. The sequence of record times is defined as follows: $T_1 = 1$ with probability 1 and, $T_n = \min\{j : X_j > X_{T_{n-1}}\}$ for $n \geq 2$. A sequence of upper record values is defined by $X_{U(n)} = X_{T_n}$, $n = 1, 2, \dots$. For details on record values and other interesting topics related to records see [Ahsanullah \(1995\)](#) and [Arnold et al. \(1998\)](#).

In this paper the estimation of the scale parameter in two exponential distributions based on record values is studied. When two record samples from exponential distributions are available, the question

¹M. Karimi : mkarimiz68@yahoo.com

of whether to pool or not to pool these two record samples is often determined via a preliminary test. If the test is not statistically significant, the pooled estimator is used; otherwise, the likelihood estimator is used. The optimum level of statistical significance for the usual preliminary test estimator are obtained in Section 2 by using the minimax regret criterion. A shrinkage estimator is also considered. The optimum values of shrinkage coefficients for the shrinkage estimator are obtained in Section 3. The proposed estimators are illustrated using an example.

2 Preliminary test estimation

Let X_1, X_2, \dots and Y_1, Y_2, \dots be sequences of independent and identically distributed random variables from two exponential models with the following probability density function:

$$f_X(x) = \frac{1}{\theta_1} \exp\left(-\frac{x}{\theta_1}\right) \quad x > 0, \quad f_Y(y) = \frac{1}{\theta_2} \exp\left(-\frac{y}{\theta_2}\right) \quad y > 0.$$

Also, suppose that we observe n_1 and n_2 upper record values from these two sequences as $\mathbf{X} = (X_{U(1)}, \dots, X_{U(n_1)})$ and $\mathbf{Y} = (Y_{U(1)}, \dots, Y_{U(n_2)})$. Therefore, the maximum likelihood estimations (MLE) of θ_1 and θ_2 are

$$\hat{\theta}_1 = \frac{1}{n_1} X_{U(n_1)}, \quad \hat{\theta}_2 = \frac{1}{n_2} Y_{U(n_2)},$$

and $\frac{2n_i \hat{\theta}_i}{\theta_i}$ has a chi-square distribution with $2n_i$, $i = 1, 2$ degrees of freedom.

Assume that a prior information about the scale parameters is $\theta_1 = \theta_2$. Then the pooled estimator for θ_1 is

$$\hat{\theta}_p = \frac{n_1 \hat{\theta}_1 + n_2 \hat{\theta}_2}{n_1 + n_2}. \quad (2.1)$$

Now consider testing $H_0 : \theta_1 = \theta_2$ against $H_1 : \theta_1 \neq \theta_2$. If the null hypothesis is not rejected we use the pooled samples for estimating the scale parameter, but if the null hypothesis is rejected then we use the individual sample for estimating the parameter.

The likelihood ratio test reject H_0 when $\frac{\hat{\theta}_1}{\hat{\theta}_2} < c_1$ or $\frac{\hat{\theta}_1}{\hat{\theta}_2} > c_2$ where $c_1 = F_{(2n_1, 2n_2), \alpha/2}$, $c_2 = F_{(2n_1, 2n_2), 1-\alpha/2}$. Therefore, a preliminary test estimator for θ_1 may be obtained as follows

$$\hat{\theta}_{pt} = \begin{cases} \hat{\theta}_p & c_1 < \frac{\hat{\theta}_1}{\hat{\theta}_2} < c_2 \\ \hat{\theta}_1 & \text{otherwise.} \end{cases} \quad (2.2)$$

The preliminary test estimator always depends on the significance level (α) of the preliminary test. Hirano (1977) applied AIC (Akaike, 1998) to determine the optimal level of significance for the preliminary test.

Lemma 2.1. *If we consider the weighted square error loss function $L(d; \theta) = \frac{(d-\theta)^2}{\theta^2}$ and $\delta = \frac{\theta_2}{\theta_1}$ then the Risk($\theta_1, \hat{\theta}_{pt}$) is:*

$$\begin{aligned} R(\delta, \alpha) &= Risk(\theta_1, \hat{\theta}_{pt}) \\ &= \delta^2 \left[\lambda^2 \frac{(n_2 + 1)}{n_2} [I_{d_2}(n_1, n_2 + 2) - I_{d_1}(n_1, n_2 + 2)] \right] \\ &\quad + \delta \left[2(\lambda - \lambda^2) [I_{d_2}(n_1 + 1, n_2 + 1) - I_{d_1}(n_1 + 1, n_2 + 1)] \right. \\ &\quad \left. - 2\lambda [I_{d_2}(n_1, n_2 + 1) - I_{d_1}(n_1, n_2 + 1)] \right] \\ &\quad + \left[\frac{1}{n_1} + (\lambda^2 - 2\lambda) \frac{(n_1 + 1)}{n_1} [I_{d_2}(n_1 + 2, n_2) - I_{d_1}(n_1 + 2, n_2)] \right] \\ &\quad + 2\lambda [I_{d_2}(n_1 + 1, n_2) - I_{d_1}(n_1 + 1, n_2)], \end{aligned}$$

The proof is straightforward.

Notice that the risk function depends on α through c_1 and c_2 . If $\delta \rightarrow 0$ or ∞ then $R(\delta, \alpha)$ converges to $R(\delta, 1)$ which is the risk of $\hat{\theta}_1$. The general shapes of $R(\delta, \alpha)$ for fixed values of n_1, n_2 and some α are in Figure 1.

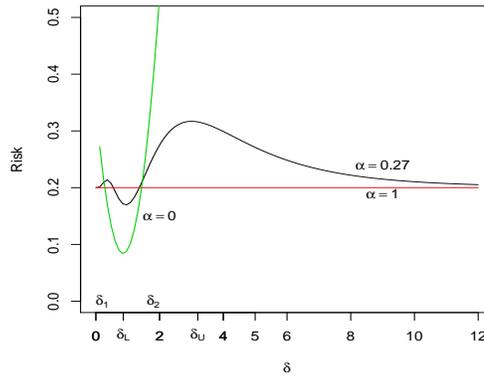


Figure 1: The risk function for some α with $n_1 = 5$ and $n_2 = 6$.

An optimal value of α is $\alpha = 1$ if $\delta \leq \delta_1$ or $\delta \geq \delta_2$ and $\alpha = 0$ otherwise, where δ_1 and δ_2 are intersections of $R(\delta, 0) = (n_1 + n_2\delta^2 + n_2^2\delta^2 + n_2^2 - 2n_2^2\delta)/(n_1 + n_2)^2$ and $R(\delta, 1) = 1/n$.

The aim is to find the optimum values of α , according to the minimax regret criterion. Since δ is unknown we seek an optimal value $\alpha = \alpha^*$ which gives a reasonable risk for all values of δ . The regret function is defined as

$$Reg(\delta, \alpha) = R(\delta, \alpha) - \inf_{\alpha} R(\delta, \alpha),$$

where $\inf_{\alpha} R(\delta, \alpha) = R(\delta, 0)$ for $\delta_1 < \delta < \delta_2$, and $\inf_{\alpha} R(\delta, \alpha) = R(\delta, 1)$, otherwise.

In such a situation, we are often able to arrive at reasonable results by minimizing the maximum regret instead of the maximum risk. The minimax regret criterion determines α^* such that

$$\sup_{\delta} \{Reg(\delta, \alpha^*)\} \leq \sup_{\delta} \{Reg(\delta, \alpha)\},$$

for every significance level $\alpha \neq \alpha^*$. From Figure 1, we can find that $Reg(\delta, \alpha)$ takes a maximum value at δ_L for $\delta \leq \delta_2$, and it takes a maximum value at δ_U for $\delta > \delta_2$. Thus the minimax regret criterion determines α^* such that $Reg(\delta_L, \alpha^*) = Reg(\delta_U, \alpha^*)$. We found numerically the optimum significance level α^* for some n_1 and n_2 . The results are given in Table 2.

3 Shrinkage estimator

The preliminary test estimator given in Section 2 uses pooled estimator $\hat{\theta}_p$ when preliminary test accepts the null hypothesis. Instead of using $\hat{\theta}_p$, we can use a linear combination of $\hat{\theta}_p$ and $\hat{\theta}_1$ when the preliminary test accepts H_0 , this gives a preliminary test shrinkage estimator which is smoother than the usual preliminary test estimator. The shrinkage estimator performs better than the usual estimator when the our guess is approximately true. Shrinkage estimators are usually defined as estimators obtained through modification of the usual (maximum likelihood, minimum variance unbiased, least squares, etc.) estimator in order to optimize some desirable criterion function.

In this section, we study the preliminary test shrinkage estimator $\hat{\theta}_S$ following the same estimation procedure by Inada (1984):

$$\hat{\theta}_S = \begin{cases} K\hat{\theta}_p + (1 - K)\hat{\theta}_1 & c_1 < \frac{\hat{\theta}_1}{\theta_1} < c_2 \\ \hat{\theta}_1 & \text{otherwise.} \end{cases} \quad (3.1)$$

If $K = 1$, $\hat{\theta}_S$ reduce to $\hat{\theta}_{pt}$. The shrinkage coefficient K is not defined explicitly as a function of the test statistic. The weighting function approach is intuitively appearing, but the mean square error of the resulting estimator usually cannot be derive unless the weighting function is in some simple form. Note that $\hat{\theta}_{pt}$ approaches θ_0 as $\alpha \rightarrow 0$ and it approaches $\hat{\theta}_{ML}$ as $\alpha \rightarrow 1$, however $\hat{\theta}_S$ approaches $\hat{\theta}_{pt}$ as $K \rightarrow 1$ and it approaches $\hat{\theta}_{ML}$ as $K \rightarrow 0$. Unfortunately, different value of significance level (α) or different value of shrinkage coefficient (K) induces a different estimator. the choice of these values depends on the decision criterion.

Lemma 3.1. Under the weighted square error loss function $L(d; \theta) = \frac{(d-\theta)^2}{\theta^2}$, the Risk($\theta_1, \hat{\theta}_S$) is:

$$R(\delta, K, \alpha) = Risk(\theta_1, \hat{\theta}_S)$$

$$\begin{aligned}
 &= \delta^2 \left[K^2 \lambda^2 \frac{(n_2 + 1)}{n_2} [I_{d_2}(n_1, n_2 + 2) - I_{d_1}(n_1, n_2 + 2)] \right] \\
 &+ \delta [2(K\lambda - K^2\lambda^2)[I_{d_2}(n_1 + 1, n_2 + 1) - I_{d_1}(n_1 + 1, n_2 + 1)] \\
 &- 2K\lambda[I_{d_2}(n_1, n_2 + 1) - I_{d_1}(n_1, n_2 + 1)]] \\
 &+ \left[\frac{1}{n_1} + (K^2\lambda^2 - 2K\lambda) \frac{(n_1 + 1)}{n_1} [I_{d_2}(n_1 + 2, n_2) - I_{d_1}(n_1 + 2, n_2)] \right] \\
 &+ 2K\lambda[I_{d_2}(n_1 + 1, n_2) - I_{d_1}(n_1 + 1, n_2)].
 \end{aligned}$$

The proof is straightforward.
 The regret function is defined as:

$$\text{Reg}(\delta, K, \alpha) = R(\delta, K, \alpha) - \inf_K R(\delta, K, \alpha).$$

Since $R(\delta, K, \alpha)$ has quadratic form w.r.t K , so it has minimum in the interval $[0,1]$. Let $R(\delta, K, \alpha) = K^2 h_2(\delta) + K h_1(\delta) + h_0(\delta)$ then $\frac{\partial R}{\partial K} = 0$ imply that $K_0 = \frac{-h_1(\delta)}{2h_2(\delta)}$ is an extremum point. Therefore

$$\inf_K \text{Risk}(\delta, K, \alpha) = \begin{cases} \min\{R(\delta, 0, \alpha), R(\delta, 1, \alpha), R(\delta, K_0, \alpha)\}, & \text{if } K_0 \in (0, 1) \\ \min\{R(\delta, 0, \alpha), R(\delta, 1, \alpha)\}, & \text{otherwise.} \end{cases}$$

After rather extensive numerical investigation the values K which attain the $\inf_K \sup_\delta \text{Reg}(\delta, K, \alpha)$ are obtained. We plot the risk function for $n_1 = 5, n_2 = 6, \alpha = 0.16$, for $K = 0, 1, 0.21$ in Figure 2.

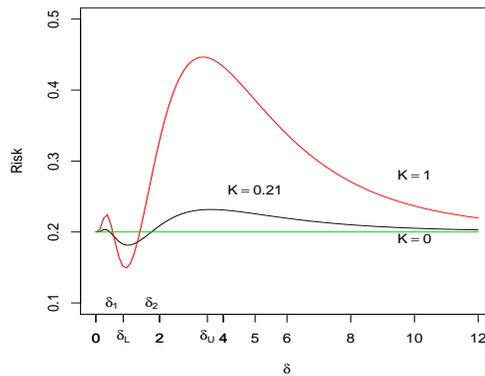


Figure 2: The risk function for some K with $n_1 = 5$ and $n_2 = 6$.

Let δ_1 and δ_2 be intersections of $\text{Risk}(\delta, 0, \alpha)$ and $\text{Risk}(\delta, 1, \alpha)$. For $\delta \leq \delta_2$, $\text{Reg}(\delta, K, \alpha)$ takes a maximum value at δ_L and for $\delta > \delta_2$, it takes a maximum value at δ_U , the reasonable estimator is

obtained when these two maximum values be equal. Thus the minimax regret criterion determines K^* such that:

$$Reg(\delta_L, K^*) = Reg(\delta_U, K^*).$$

To find the optimal value of K , two cases are considered for α :

Case I: Let $\alpha = 0.16$, that is the AIC optimal level of significance (see ?), which is independent of n . Table 1 presents the values of K^* for some n_1 and n_2 .

Case II: Let $\alpha = \alpha^*$ equal the significance level of pre-test. Table 2 presents the values of K^* in this case for some n_1 and n_2 .

Table 1: Optimal value K for some n_1 and n_2 with $\alpha = 0.16$.

	n_1					
n_2	2	3	4	5	7	10
2	0.17	0.23	0.29	0.32	0.38	0.42
3	0.14	0.19	0.24	0.27	0.32	0.36
4	0.12	0.17	0.21	0.24	0.29	0.33
5	0.11	0.16	0.20	0.22	0.27	0.31
7	0.10	0.14	0.18	0.20	0.24	0.28
10	0.10	0.13	0.17	0.19	0.23	0.26

Table 2: Optimal value K for some n_1 and n_2 with optimal $\alpha = \alpha^*$.

	$n_1 = 2$		$n_1 = 3$		$n_1 = 4$		$n_1 = 5$		$n_1 = 7$		$n_1 = 10$	
n_2	α^*	K^*										
2	0.38	0.21	0.30	0.29	0.27	0.34	0.24	0.37	0.22	0.42	0.20	0.45
3	0.42	0.15	0.34	0.23	0.29	0.28	0.27	0.31	0.24	0.35	0.21	0.39
4	0.44	0.12	0.36	0.19	0.31	0.24	0.28	0.27	0.25	0.32	0.22	0.35
5	0.46	0.11	0.38	0.17	0.33	0.22	0.30	0.25	0.26	0.29	0.23	0.33
7	0.49	0.10	0.40	0.15	0.35	0.19	0.32	0.22	0.28	0.26	0.25	0.30
10	0.51	0.09	0.42	0.14	0.37	0.17	0.33	0.20	0.29	0.24	0.26	0.28

Remark 3.2. For the general case, if we observe upper record values $X_{U(1)}, \dots, X_{U(n_1)}$ and $Y_{U(1)}, \dots, Y_{U(n_2)}$ from location-scale exponential distribution with the following probability density function

$$f_X(x) = \frac{1}{\theta_1} \exp\left(-\frac{x - \eta_1}{\theta_1}\right) \quad x > \eta_1, \quad f_Y(y) = \frac{1}{\theta_2} \exp\left(-\frac{y - \eta_2}{\theta_2}\right) \quad y > \eta_2,$$

the MLE's of θ_1 and θ_2 are

$$\hat{\theta}_1 = \frac{1}{n_1}(X_{U(n_1)} - X_{U(1)}), \quad \hat{\theta}_2 = \frac{1}{n_2}(Y_{U(n_2)} - Y_{U(1)}),$$

and $2n_i\hat{\theta}_i/\theta_i$ has a chi-square distribution with $2n_i - 2$, $i = 1, 2$ degrees of freedom. In this case, we have the same result as the special case in absence of location parameter with a new statistic and degree of freedom.

4 Numerical study

In this section, we first illustrate the proposed estimators using an example. Then, these estimators are compared using simulation.

4.1 An example

The following the simulated record values are given by ?. The first record values which are generated from location-scale exponential distribution with $\theta_1 = 1$ and $\eta_1 = 3$ are 3.105, 6.158, 6.296, 6.824, 7.282, 10.200, 10.240, and 11.669, and first record values which are generated from location-scale exponential distribution with $\theta_2 = 1$ and $\eta_2 = 1$ are 1.177, 2.430, 4.090, 4.349, 4.624, 5.655, 6.021, and 6.987.

The MLE's of θ_1 and θ_2 are $\hat{\theta}_1 = 1.0705$ and $\hat{\theta}_2 = 0.7262$. Since the hypothesis $H_0 : \theta_1 = \theta_2$ is not rejected the preliminary test and optimal shrinkage estimations are $\hat{\theta}_{pt} = 0.8984$, and $\hat{\theta}_S = 1.0292$, respectively. It can be seen that the optimal shrinkage estimation is very close to the true value $\theta_1 = 1$.

References

- Ahsanullah, M. (1995). *Record statistics*, Nova Science Publishers Commack, New York.
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle, *Selected Papers of Hirotugu Akaike*, Springer Series in Statistics, New York, pp. 199213.
- Arnold, B., Balakrishnan, N., and Nagaraja, H. (1998). *Records*, John Wiley and Sons, New York.
- Baklizi, A. (2013). Interval estimation of the stress-strength reliability in the two-parameter exponential distribution based on records, *Journal of Statistical Computation and Simulation*, pp. 110.
- Hirano, K. (1977). Estimation procedures based on preliminary test, shrinkage technique and information criterion, *Annals of the Institute of Statistical Mathematics*, **29(1)**, 21–34.

- Inada, K. (1984). A minimax regret estimator of a normal mean after preliminary test, *Annals of the Institute of Statistical Mathematics*, **36(1)**, 207–215.
- Prakash, G. and Singh, D. (2008). Shrinkage estimation in exponential type-II censored data under LINEX loss, *Journal of the Korean Statistical Society*, **37(1)**, 53–61.
- Toyoda, T. and Wallace, T. (1975). Estimation of variance after a preliminary test of homogeneity and optimal levels of significance for the pre-test, *Journal of Econometrics*, **3(4)**, 395–404.



Rate of Convergence for Sample Covariance Operator of H -valued Periodically Correlated Time Series

Atefeh Zamani,¹

¹Department of Statistics, Faculty of Science, Shiraz University, Shiraz 7146713565,
IRAN.

Abstract: Periodically correlated (PC) or cyclostationary time series, as a special class of nonstationary time series, is of great importance in climatology, hydrology, electrical engineering, signal processing and economics. These time series have a periodic structure in their means and covariances. Consequently, an important feature in studying PC time series is identifying the value of the period, T . This paper focuses on convergence rate of covariance operators of Hilbertian valued periodically correlated processes, which can be applied while determining the value of T in these time series.

Keywords H -valued Periodically Correlated Processes, Time Series, Convergence Rate.

Mathematics Subject Classification (2010): 62M10 60B12.

1 Introduction

Recently, a new aspect of statistical data analysis, called functional data analysis, attracts the attention of several researchers. Developments in probability theory and statistical methods, along with the progress of technologies, which allow us to store more and more information, have led to necessity to pass from describing numbers or a finite set of numbers to describing infinite dimensional random variables, random fields, random sets, random functions, etc. Studying infinite dimensional random variables is certainly of great importance, since despite of being a heavy challenge, it combines original theoretical and practical problems, which can be applied while developing new methodologies in other fields.

Time series analysis, as an important method in analyzing data, has been extended to infinite dimensional random processes and various models have been introduced in this field (see for example [Bosq \(2000\)](#)). Periodically correlated random processes, with periodic behavior in mean and covariance, is a natural extension of stationary processes, [Gladyshev \(1961\)](#). The infinite dimensional periodically correlated time series is scrutinized by several researchers, such as [Hurd \(1989\)](#) and [Soltani and Hashemi \(2011\)](#). This paper tries to study the asymptotic behavior of mean and covariance operator, which may be applied while determining the value of the period, T , in such processes.

¹A. Zamani: zamania@shirazu.ac.ir

2 Preliminaries

This paper is concern with random processes with values in Hilbert spaces. Consequently, various form of spaces, inner products and norms need to be introduced. This section is devoted to preliminary notations and definitions, which will be used in the subsequent sections.

Let (Ω, \mathcal{F}, P) be a probability space and $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ stands for the Hilbert space of all complex-valued random variables with finite second moment on (Ω, \mathcal{F}, P) . The inner product in this Hilbert space is defined as $E(X\bar{Y})$, where E stands for the expectation. Moreover, consider H to be a separable Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle_H$. The space of bounded linear operators on H is defined as $\mathcal{L}(H) = \left\{ \mathbf{A} : H \rightarrow H : \|\mathbf{A}\| = \sup_{x \in H, x \neq 0} \frac{\|\mathbf{A}x\|_H}{\|x\|_H} < \infty \right\}$.

Let $\xi : \Omega \rightarrow H$ be a random variable with values in H ; i.e., $\xi_t : \Omega \rightarrow H$ is \mathcal{F}/\mathcal{B} measurable, in which \mathcal{B} is the Borel field in H . The random variable ξ is called an H -valued random variable and the sequence $\{\xi_t, t \in \mathbb{Z}\}$, where \mathbb{Z} is the set of integers, is called an H -valued discrete time stochastic process. The process $\{\xi_t, t \in \mathbb{Z}\}$ is called and strongly second order (SSO in abbreviation) if $\|\xi_t\| \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. For the H -valued SSO random variable ξ the expectation and covariance operators are introduced in terms of Bochner integral and tensorial product as follows:

Definition 2.1 (Expectation and Covariance Operator). *Let ξ be an H -valued SSO random variable. The expectation of ξ , denoted by $\mathbb{E}\xi$, is defined as the Bochner integral of ξ over Ω ; i.e., $\mu = \mathbb{E}\xi := B - \int_{\Omega} \xi dP$. Equivalently, the covariance operator of the random variables ξ and η is an operator given by $\mathbf{C}_{\xi, \eta} := \mathbb{E}(\xi - \mathbb{E}\xi) \otimes (\eta - \mathbb{E}\eta)$, where \otimes stands for the tensorial product and is defined as $(\xi \otimes \eta)z = \overline{\langle \xi, z \rangle_H} \eta$, for $\forall z \in H$.*

Based on the given definitions and notations, an H -valued periodically correlated processes can be defined. An H -valued SSO stochastic process is called periodically correlated (PC in abbreviation) if there exists an integer $T > 0$ such that for every $t \in \mathbb{Z}$:

$$\mu_t := \mathbb{E}(\xi_t) = \mathbb{E}(\xi_{t+T}), \quad (2.1)$$

and

$$C_{t, t+\tau} := \mathbb{E}\{(\xi_t - \mu_t) \otimes (\xi_{t+\tau} - \mu_{t+\tau})\} = \mathbb{E}\{(\xi_{t+T} - \mu_{t+T}) \otimes (\xi_{t+T+\tau} - \mu_{t+T+\tau})\}. \quad (2.2)$$

The smallest T , for which (2.1) and (2.2) hold, is called the period of the process. If $T = 1$, then the process is called stationary.

Consider the periodically correlated time series $\boldsymbol{\xi} = \{\xi_1, \xi_2, \dots, \xi_n\}$. The first step in analyzing the time series $\boldsymbol{\xi}$ is determining the value of the period T . Since in periodically correlated time series the mean and the covariance have a common value as the period, this paper focuses on ascertaining the

convergence rate of covariance operator. For this purpose, the sample autocovariance operator, which is defined by

$$\hat{R}(\tau) := \frac{1}{n-\tau} \sum_{t=1}^{n-\tau} (\xi_t - \bar{\xi}_n) \otimes (\xi_{t+\tau} - \bar{\xi}_n), \tag{2.3}$$

is studied in more details. Note that $\bar{\xi}_n = \frac{1}{n} \sum_{t=1}^n \xi_t$.

Lemma 2.2. *Let ξ be a sequence of H -valued periodically correlated process with period T , for which $\sum_{s=0}^{\infty} \sum_{t=1}^T \|C_{s,t}\|_0 < \infty$. Then,*

$$\lim_{n \rightarrow \infty} \left\| \mathbb{E}(\bar{\xi}_n) - \bar{\mu} \right\|_H = 0, \tag{2.4}$$

$$\lim_{n \rightarrow \infty} \left\| \mathbb{E}(\hat{R}(\tau)) - R_\mu(\tau) - \bar{R}(\tau) \right\| = 0, \tag{2.5}$$

where

$$\begin{aligned} \bar{\mu} &= \frac{1}{T} \sum_{t=1}^T \mu_t \\ R_\mu(\tau) &= \frac{1}{T} \sum_{t=1}^T (\mu_t - \bar{\mu}) \otimes (\mu_{t+\tau} - \bar{\mu}) \\ \bar{R}(\tau) &= \frac{1}{T} \sum_{t=1}^T C_{t,t+\tau}, \end{aligned}$$

Proof. Let $n = kT + l$, then:

$$\begin{aligned} \lim_{n \rightarrow \infty} \left\| \mathbb{E}(\bar{\xi}_n) - \frac{1}{T} \sum_{t=1}^T \mathbb{E}(\xi_t) \right\|_H &= \lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{t=1}^n \mathbb{E}(\xi_t) - \frac{1}{T} \sum_{t=1}^T \mathbb{E}(\xi_t) \right\|_H \\ &= \lim_{k \rightarrow \infty} \frac{1}{T(kT+l)} \left\| (T-l) \mathbb{E}(\xi_1) + \dots + (T-l) \mathbb{E}(\xi_l) \right. \\ &\quad \left. - l \mathbb{E}(\xi_{l+1}) - \dots - l \mathbb{E}(\xi_T) \right\|_H = 0. \end{aligned}$$

Moreover,

$$\begin{aligned} &\lim_{n \rightarrow \infty} \left\| \mathbb{E}(\hat{R}(\tau)) - R_\mu(\tau) - \bar{R}(\tau) \right\| \\ &= \lim_{n \rightarrow \infty} \left\| \frac{1}{n-\tau} \sum_{t=1}^{n-\tau} \mathbb{E} \left\{ (\xi_t - \mu_t) \otimes (\xi_{t+\tau} - \mu_{t+\tau}) + (\mu_t - \bar{\xi}_n) \otimes (\xi_{t+\tau} - \mu_{t+\tau}) \right. \right. \\ &\quad \left. \left. + (\xi_t - \mu_t) \otimes (\mu_{t+\tau} - \bar{\xi}_n) + (\mu_t - \bar{\xi}_n) \otimes (\mu_{t+\tau} - \bar{\xi}_n) \right\} - R_\mu(\tau) - \bar{R}(\tau) \right\| \end{aligned}$$

$$\begin{aligned}
 &= \lim_{n \rightarrow \infty} \left\| \frac{1}{n - \tau} \sum_{t=1}^{n-\tau} \{ C_{t,t+\tau} + \mathbb{E} \{ (\mu_t - \bar{\mu}_n + \bar{\mu}_n - \bar{\xi}_n) \otimes (\xi_{t+\tau} - \mu_{t+\tau}) \right. \\
 &\quad + (\xi_t - \mu_t) \otimes (\mu_{t+\tau} - \bar{\mu}_n + \bar{\mu}_n - \bar{\xi}_n) \\
 &\quad \left. + (\mu_t - \bar{\mu}_n + \bar{\mu}_n - \bar{\xi}_n) \otimes (\mu_{t+\tau} - \bar{\mu}_n + \bar{\mu}_n - \bar{\xi}_n) \} \right\} - R_\mu(\tau) - \bar{R}(\tau) \Big\| \\
 &\leq \lim_{n \rightarrow \infty} \left\| \frac{1}{n - \tau} \sum_{t=1}^{n-\tau} C_{t,t+\tau} - \bar{R}(\tau) \right\| \\
 &\quad + \lim_{n \rightarrow \infty} \left\| \frac{1}{n^2(n - \tau)} \left\{ (n - \tau) \sum_{t'=n-\tau+1}^n \sum_{t''=n-\tau+1}^n C_{t,t''} - \tau \sum_{t=n-\tau+1}^n \sum_{t'=\tau+2}^n C_{t',t''} \right. \right. \\
 &\quad \left. \left. - \tau \sum_{t=1}^{n-\tau} \sum_{t'=1}^{\tau+1} C_{t,t'} - (n + \tau) \sum_{t=1}^{n-\tau} \sum_{t'=\tau+2}^n C_{t',t''} \right\} \right\| \\
 &\quad + \lim_{n \rightarrow \infty} \left\| \frac{1}{n - \tau} \sum_{t=1}^{n-\tau} (\mu_t - \bar{\mu}_n) \otimes (\mu_{t+\tau} - \bar{\mu}_n) - R_\mu(\tau) \right\| = 0.
 \end{aligned}$$

□

As can be seen, the expected value of the sample autocovariance operator can be approximated by the sum of two terms, covariance operator of the periodic mean function μ_t (at lag τ), $R_\mu(\tau)$, and the average of the periodic autocovariance operators $C_{t,t+\tau}$, $t = 1, 2, \dots, T$, $\bar{R}(\tau)$. It can easily be seen that $R_\mu(\tau)$ is periodic in τ with the same period as μ_t . However, the behavior of the second term should be studied more accurately. Consequently, studying the behavior of the sample autocovariances operator might be of significance for indicating the periodic properties of process ξ . For this purpose, the limiting behavior of the sample autocovariances operator will be scrutinized and the rate of almost sure convergence will be determined.

3 Rate of Convergence in the Sample Autocovariances Operators

Consider a sequence of purely nondeterministic, full-rank $HPC - T$ processes, ξ . By Wold decomposition, this process can be presented as,

$$\xi_t = \mu_t + \sum_{j=0}^{\infty} a_j(t) \varepsilon_{t-j}, \tag{3.1}$$

where $\mathcal{I} = \{\varepsilon_t, t \in \mathbb{Z}\}$ is a set of orthonormal innovation processes, $a_j(t)$ are linear operators on H , for which $\sum_{j \geq 0} \|a_j(t)\|_0^2 < \infty$, and $a_j(t + kT) = a_j(t)$ for every j, k, t , with $j \geq 0$.

Let $\mathcal{I} = \{\varepsilon_t, t \in \mathbb{Z}\}$ be a sequence of martingale difference with respect to an increasing sequence of σ -field \mathcal{F}_t ; i.e.,

$$\mathbb{E}^{\mathcal{F}_{t-1}}(\varepsilon_t) = 0, \quad \text{a.s.} \quad \forall t \in \mathbb{Z}, \tag{3.2}$$

where 0 is the zero of H . Moreover, assume that

$$\mathbb{E}^{\mathcal{F}_t}(\varepsilon_t) = 0, \quad \text{a.s.} \quad \forall t \in \mathbb{Z}, \tag{3.3}$$

$$E^{\mathcal{F}_{t-1}} \|\varepsilon_t\|^2 \leq d_t, \quad \text{a.s.} \quad \forall t \in \mathbb{Z}, \tag{3.4}$$

$$\sup_t E \|\varepsilon_t\|^\lambda \leq \infty, \quad \text{for some } \forall \lambda > 4. \tag{3.5}$$

Theorem 3.1. *For some $\delta > 0$ and $1 \leq G(n) = o(n)$,*

$$\max_{1 \leq \tau \leq G(n)} \left\| \hat{R}(\tau) - (R_\mu(\tau) + \bar{R}(\tau)) \right\| = o\left(n^{-1/2} (G(n) \log n)^{2/\lambda} (\log \log n)^{2(1+\delta)/\lambda}\right) \tag{3.6}$$

This theorem can be proved using the following lemma, which is a direct result of [Tian \(1988\)](#).

Lemma 3.2. *Let $\{\eta_i, i = 1, 2, \dots\}$ be a sequence of H -valued stochastic processes, where positive constants C, ν and a positive sequence a_n exist, such that $\limsup a_{2n}/a_n < \infty$ and for all $n, x > 0$, $P(\max_{1 \leq i \leq n} \|\eta_i\|_H > x) < Ca_n x^{-\nu}$. Then, for any $\delta > 0$,*

$$\|\eta_n\|_H = o\left(\left(a_n (\log n) (\log \log n)^{1+\delta}\right)^{1/\nu}\right). \tag{3.7}$$

Proof of Theorem 3.1. Denote the periodically sampled autocovariance operator of lag τ and with starting point i as:

$$\hat{R}(i, \tau) = (h_n + 1)^{-1} \sum_{m=0}^{h_n} (\xi_{i+mT} - \bar{\xi}_{n-\tau}(i, T)) \otimes (\xi_{i+\tau+mT} - \bar{\xi}_n(i + \tau, T)), \tag{3.8}$$

where

$$\bar{\xi}_n(i, T) = (h(n - i, T) + 1)^{-1} \sum_{j=0}^{h(n-i, T)} \xi_{i+jT}, \quad 1 \leq i, T < n, \tag{3.9}$$

and $h_n = h(n - \tau - i, T) = \lfloor \frac{n-\tau-i}{T} \rfloor$. The relation presented in (3.8) can be rewritten as:

$$\begin{aligned} \hat{R}(i, \tau) &= (h_n + 1)^{-1} \sum_{m=0}^{h_n} (\xi_{i+mT} \otimes \xi_{i+\tau+mT}) - \left[\bar{\xi}_{n-\tau}(i, T) \otimes \left((h_n + 1)^{-1} \sum_{m=0}^{h_n} \xi_{i+\tau+mT} \right) \right] \\ &\quad - \left[\left((h_n + 1)^{-1} \sum_{m=0}^{h_n} \xi_{i+mT} \right) \otimes \bar{\xi}_n(i + \tau, T) \right] + (h_n + 1)^{-1} \sum_{m=0}^{h_n} \bar{\xi}_{n-\tau}(i, T) \otimes \bar{\xi}_n(i + \tau, T) \end{aligned}$$

$$= (h_n + 1)^{-1} \sum_{m=0}^{h_n} (\xi_{i+mT} \otimes \xi_{i+\tau+mT}) - \bar{\xi}_{n-\tau}(i, T) \otimes \bar{\xi}_n(i + \tau, T).$$

Besides, $\hat{R}(\tau)$ can be presented in terms of $\hat{R}(i, \tau)$ as follows:

$$\begin{aligned} \hat{R}(\tau) &= \frac{1}{n - \tau} \sum_{i=1}^T \sum_{m=0}^{h_n} (\xi_{i+mT} - \bar{\xi}_n) \otimes (\xi_{i+\tau+mT} - \bar{\xi}_n) \\ &= \frac{1}{n - \tau} \left[\sum_{i=1}^T (h_n + 1) \left(\hat{R}(i, \tau) + \bar{\xi}_{n-\tau}(i, T) \otimes \bar{\xi}_n(i + \tau, T) \right) \right. \\ &\quad \left. - \sum_{i=1}^T \sum_{m=0}^{h_n} \bar{\xi}_n \otimes \xi_{i+\tau+mT} - \sum_{i=1}^T \sum_{m=0}^{h_n} \xi_{i+mT} \otimes \bar{\xi}_n + \sum_{i=1}^T \sum_{m=0}^{h_n} \bar{\xi}_n \otimes \bar{\xi}_n \right] \\ &= \sum_{i=1}^T \frac{(h_n + 1)}{n - \tau} \hat{R}(i, \tau) + \frac{(h_n + 1)}{n - \tau} \sum_{i=1}^T (\bar{\xi}_{n-\tau}(i, T) - \bar{\xi}_n) \otimes (\bar{\xi}_n(i + \tau, T) - \bar{\xi}_n). \end{aligned}$$

Note that $\frac{h_n+1}{n-\tau} \rightarrow \frac{1}{T}$, as $n \rightarrow \infty$, and, consequently, to prove (3.6) it suffices to show that

$$\left\| \frac{1}{n - \tau} \sum_{m=0}^{h_n} (\xi_{i+mT} - \mu_i) \otimes (\xi_{i+\tau+mT} - \mu_{i+\tau}) - \frac{1}{T} C_{i, i+\tau} \right\| = o(1) \tag{3.10}$$

$$\max_{i \leq i \leq G(n)} \left\| \bar{\xi}_{n-\tau}(i, T) - \mu_i \right\|_H = o(\alpha(n)) \tag{3.11}$$

$$\max_{i \leq i \leq G(n)} \left\| \frac{(h_n + 1)}{n - \tau} \sum_{i=1}^T (\bar{\xi}_{n-\tau}(i, T) - \bar{\xi}_n) \otimes (\bar{\xi}_n(i + \tau, T) - \bar{\xi}_n) - R_\mu(\tau) \right\| = o(\alpha(n)), \tag{3.12}$$

almost surely and where $\alpha(n) = n^{-1/2} (G(n) \log n)^{2/\lambda} (\log \log n)^{2(1+\delta)/\lambda}$.

The equality (3.10) can be proved easily using the continuity of tensorial products. Therefore, in the first step, (3.11) will be proved. Set $\zeta_t = \xi_t - \mu_t$.

$$S(n, i) = (h(n - i, T) + 1) \bar{\zeta}_n(i, T)$$

$$= \sum_{k \geq 0: i+jT-k \in D^+} \sum_{j=0}^{h(n-i, T)} a_k(i) \varepsilon_{i+jT-k}$$

Define $\varepsilon_m(i, k) := \sum_{j=0}^m \varepsilon_{i+jT-k}$. It is easy to show that $\varepsilon_m(i, k)$ is a martingale and, therefore, using Doob' inequality and Burkholder's inequality we have:

$$\left(E \left(\max_{1 \leq m \leq M} \|\varepsilon_m(i, k)\|_H \right)^\lambda \right)^{1/\lambda} \leq \frac{\lambda}{1 - \lambda} \max_{1 \leq m \leq M} \left(E \|\varepsilon_m(i, k)\|_H^\lambda \right)^{1/\lambda}$$

$$\begin{aligned} &\leq E \left(\sum_{m=1}^M \|\varepsilon_{i+jT-k}\|_H^2 \right)^{1/2} \\ &\leq CM^{1/2}. \end{aligned} \tag{3.13}$$

By Wold decomposition, (3.5) and (3.13), it can be concluded that

$$E \max_{1 \leq m \leq M} \|S(n, i)\|_H^\lambda \leq CN^{1/2},$$

where N depends on $\sum_{j \geq 0} \|a_j(t)\|^2$. By the Markov inequality for H -valued random processes, we have:

$$P \left(\max_{1 \leq m \leq M} \|S(n, i)\|_H \geq x \right) \leq CN^{\lambda/2} x^{-\lambda},$$

and, consequently,

$$P \left(\max_{1 \leq i \leq G(N)} \max_{1 \leq m \leq M} \|S(n, i)\|_H \geq x \right) \leq CG(N) N^{\lambda/2} x^{-\lambda}.$$

Using Lemma 3.2, we have,

$$\max_{1 \leq i \leq G(N)} \|S(n, i)\|_H = o \left(n^{-1/2} (G(n) \log n)^{1/\lambda} (\log \log n)^{(1+\delta)/\lambda} \right),$$

which proves (3.5). The same argument, along with continuity of tensorial product, results in (3.12). Therefore, the proof is completed. □

Since the mean and covariance operator of H -valued periodically correlated process have a common value T as their period, delineating the asymptotic behavior of the sample autocovariance can be applied while determining the value of T , Tian (1988).

References

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, vol. 149. Berlin: Springer.

Gladyshev, E.G. (1961). Periodically Correlated Random Sequences, *Soviet Math. Dokl.*, **2**, 385–388.

Hurd, H.L. (1989). Representation of Strongly Harmonizable Periodically Correlated Processes and Their Covariance. *Journal Multivariate Analysis.* **29**, 53–67.

Soltani, A.R. and Hashemi, M. (2011). Periodically correlated Autoregressive Hilbertian, *Statistical Inference for Stochastic processes*, **14**, 177–188.

Stroock, D.W. (1993). *Probability Theory, An Analytic View*, Cambridge University Press.

Tian, C. J. (1988). A Limiting Property of Sample Autocovariances of Periodically Correlated Processes with Application to Period Determination. *Journal of Time Series Analysis*, **9(4)**, 411–417.



Skew Normal Autoregressive Models with Markov Property

Hamid Reza Zareifard¹, Fardin Malek Mohammadi²

Abstract: Recently some authors have introduced in the literature stationary stochastic processes, in the time and in the spatial domains, whose finite-dimensional marginal distributions are multivariate skew-normal (SN). In this work, we study the construction of skew normal autoregressive (AR) models proceeded by replacing the normal noise by a skew normal noise. The focus of the paper is to introduce a skewed AR model with Markov property. In this line, we show that AR models driven by closed skew-normal (CSN) noise enjoy some of the useful properties of Gaussian AR models.

Keywords Autoregressive models, Markov property, Closed skew-normal distribution.

1 Introduction

The most commonly used model for time series data is the autoregressive processes which are also Markov processes. The model for an autoregressive process of order 1 is often expressed as

$$x_t = \phi x_{t-1} + \epsilon_t, \quad |\phi| < 1, \quad (1.1)$$

where $\{\epsilon_t\}$ is Gaussian white noise with variance σ^2 . A nice feature of the AR models is conditional independence property, hence the term Markov. Conditional independence is a powerful concept in where two random variables are called conditionally independent given other variables. Assumptions about conditional independence are not stated explicitly here, but show up more clearly if we express (1.1) in the conditional form

$$x_t | x_1, \dots, x_{t-1} \sim N(\phi x_{t-1}, \sigma^2), \quad t = 2, \dots, n.$$

It is well known that the success of Gaussian AR models can be traced to the simple fact that the finite-dimensional distribution of the observed data $\mathbf{X} = (X_1, \dots, X_n)'$ is multivariate normal.

However, Gaussianity assumption for the white noise might be overly restrictive to represent the data. The real data could be highly non-Gaussian and may show features like skewness. In this setting, to reduce unrealistic assumptions for conditional autoregressive models, Prates et al. (2012) considered a non-Gaussian distribution for the random effects where incorporates skewness as well as heavy tail behavior of the data. However, although the covariance matrix of their random effects depends on its

²Speaker's Name: fardinmalek13@yahoo.com

neighbors, the Markov property does not supported. Also, for the class of autoregressive and moving average (ARMA) models, [Pourahmadi \(2007\)](#) studied the construction of stationary skew normal ARMA models for colored skew normal noise, but he did not consider the Markov property for his models. To capture skewness for multiplicative autoregressive models, [Karimi and Mohammadzadeh \(2010\)](#) proposed a closed skew normal model for spatial correlated errors. But, their model does not support the Markov property.

Given the multivariate skew-normal and related families of distributions ([Azzalini \(1985\)](#); [Azzalini and Capitanio \(1999\)](#); [Capitanio and Azzalini \(2003\)](#); [Gonzalez-Farias et al. \(2004\)](#)), it is desirable to choose the most promising class of SN distributions which is flexible enough to meet the needs of time series analysis. More precisely, our aim is to introduce a skew normal AR model with the hope of mimicking the success of Gaussian AR model as much as possible. In this line, the focus of the paper is on the property of conditional independence of the skew normal AR models. We show that this property does not hold for all classes of SN random variables, but the situation is better for AR models driven by closed skew-normal ([Gonzalez-Farias et al. \(2004\)](#)) noise. More precisely, we show that working within the family of closed skew-normal distributions \mathbf{X} remains in the CSN class as expected and the Markov property is supported.

2 Skewed AR models

In this section, we intend to define a skew normal AR model by analogy with the definition of Gaussian AR model. Our aim is to assess the suitability and wider applicability of SN distributions in AR models with the hope of mimicking the success of Gaussian AR model as much as possible. To this end, we retain (1.1), but replace ϵ_t by an asymmetric noise. In this line, we need to study the following two questions:

Q₁: How can we choose a skewed distribution for ϵ_t such that the distribution of \mathbf{X} belong to the same class as that of the \mathbf{X} ?

Q₂: What conditions are necessary to achieve the Markov property for the skewed AR model?

To emulate the success of normal autoregressive models we need to modify that of the family of multivariate asymmetric distributions. In recently years, authors have attempted to introduce different kinds of multivariate SN distributions. Using this family of distributions, subsequently, two skew normal AR models are defined and the properties of them are studied.

2.1 SN AR models

We are going to start with multivariate SN distribution (Azzalini and Capitanio (1999); Capitanio and Azzalini (2003)). This distribution that enjoys some of the useful properties of normal distribution, is denoted by $SN_n(\mu, \Sigma, \nu)$ and its density function is of the form

$$f(\mathbf{y}) = \phi_n(\mathbf{y}; \mu, \Sigma)\Phi_{(0+\nu D^{-1}(\mathbf{y} - \mu); 0, 1)}/\Phi(\nu; 0, 1); \quad \mathbf{y} \in \mathbb{R}^n,$$

where $\phi_n(\mathbf{y}; \mu, \Sigma)$, and $\Phi_n(\mathbf{y}; \mu, \Sigma)$ are the probability density function (*pdf*) and cumulative distribution function (*cdf*), respectively, of the n -dimensional normal distribution with mean vector μ and covariance matrix Σ . Also $\Sigma = DRD = (i_j)$ is positive-definite matrix and R is a correlation matrix, $D = \text{diag}(\tau_1, \dots, \tau_n)^{1/2}$, $\nu \in \mathbb{R}$ and $\delta \in \mathbb{R}^n$ is a vector of skewness parameters. Finally, $\tau_0 = \tau/\delta$ with $\delta = (1 + \delta'R)^{-1/2}$. This class clearly extends the multivariate normal distributions through the parameter vector δ which regulates the shape of the distribution, and for $\delta = \mathbf{0}$ it reduces to the class of normal distributions.

To get this distribution, Capitanio and Azzalini (2003) consider a $(n + 1)$ -dimensional normal random vector

$$\mathbf{W}^* = (W_0, W_1, \dots, W_n)' = \begin{pmatrix} W_0 \\ \mathbf{W} \end{pmatrix} \sim N_{n+1} \left(\mathbf{0}, \begin{pmatrix} 1 & \delta'R \\ \delta R & R \end{pmatrix} \right).$$

Then, the probability density function of $\mathbf{Y} = \mu + D(\mathbf{W} | W_0 > \nu)$ is of the pointed form. It can be shown that similarly normal distribution, this family is closed under marginalization and conditioning, but unlike normal distribution the joint distribution for i.i.d. SN random variables is not the multivariate SN one. It can be shown that similarly normal distribution, this family is closed under marginalization and conditioning, but unlike normal distribution the joint distribution for i.i.d. SN random variables is not the multivariate SN one.

Based on these notes, it is not a good idea to assume w_1, \dots, w_n are i.i.d. from the univariate SN family since first question (Q1) is not answered. A way of weakening the requirement of w_1, \dots, w_n being i.i.d. is to take the simple jointly SN distributed noise for the finite-dimensional distribution of $\mathbf{w} = (w_1, \dots, w_n)$. In this setting, we assume $\mathbf{w} \sim SN_n(0, 2I_n, \mathbf{1}_n, 0)$. An AR model with this definition for the noise is called *SN AR model*.

Now the density for \mathbf{X} can easily be driven from closure property of the SN distribution under linear transformations. Under this choice for the noise, the covariance matrix of \mathbf{w} is

$$Cov(\mathbf{w}) = 2 I_n - 0.637 \frac{22}{\sqrt{1 + n^2}} \mathbf{1}\mathbf{1}',$$

and so the covariance matrix of \mathbf{w} is not a diagonal matrix and it is absolutely dense. It means that even under this simple choice, we can not support the conditional independence property based on SN

distribution for the finite-dimensional distribution of \mathbf{Y} . It means under multivariate SN distribution, \mathbf{Q}_1 and \mathbf{Q}_2 are not answered simultaneously. The reason for this is that this distribution was obtained with conditioning only on one normal variable W_0 as you saw. To address both two questions, another skewed AR model is subsequently introduced.

2.2 CSN AR model

Alternatively, [Gonzalez-Farias et al. \(2004\)](#) extended SN distribution and proposed the multivariate closed skew normal distribution which is closed under marginalization, conditioning and linear combination. Also unlike SN family the joint distribution of i.i.d. CSN random variables is the multivariate CSN distribution. An n -dimensional random vector \mathbf{Y} is said to have a multivariate closed skew-normal distribution, denoted by $CSN_{n,m}(\boldsymbol{\mu}, \Sigma, \Gamma, \boldsymbol{\nu}, \Delta)$, if its density function is of the form

$$f(\mathbf{y}) = \phi_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma) \Phi_m(\Gamma(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta) / \Phi_m(0; \boldsymbol{\nu}, \Delta + \Gamma \Sigma \Gamma'), \quad (2.1)$$

where $\boldsymbol{\nu} \in \text{Re}^m$, $\Gamma \in \text{Re}^{m \times n}$ and $\Delta \in \text{Re}^{m \times m}$ is covariance matrix. To provide this distribution, [Gonzalez-Farias et al. \(2004\)](#) consider a $(n + m)$ -dimensional normal random vector

$$\mathbf{W}^* = (W_{0_1}, \dots, W_{0_m}, W_1, \dots, W_n)' = \begin{pmatrix} \mathbf{W}_0 \\ \mathbf{W} \end{pmatrix} \sim N_{n+m} \left(\mathbf{0}, \begin{pmatrix} \Delta + \Gamma \Sigma \Gamma' & \Gamma \Sigma \\ \Sigma \Gamma' & \Sigma \end{pmatrix} \right).$$

Then, the probability density function of $\mathbf{Y} = (\mathbf{W} | \mathbf{W}_0 > \boldsymbol{\nu})$ is the multivariate closed skew-normal (2.1). To increase the amount of skewness in the vector \mathbf{Y} as well as to simplify the interpretation of this density, [Allard and Naveau \(2007\)](#) assumed that $m = n$, $\Gamma = I_n$, $\boldsymbol{\nu} = 0$ and $\Delta = \Sigma$. This model is also referred as to the homotopic model.

Now working within the family of CSN distributions and with respect to homotopic model, we assume

$$W_1, \dots, W_n \stackrel{iid}{\sim} CSN_{1,1}(0, 2, \gamma, 0, 2) \Leftrightarrow \mathbf{W} \sim CSN_{n,n}(\mathbf{0}, 2I_n, I_n, \mathbf{0}, 2I_n).$$

where $\gamma = (\gamma_1, \dots, \gamma_n)$ and 2 are skewness and scale parameter, respectively. An AR model with this definition for the noise is called *CSN AR model*. In the following theorem, the properties of CSN AR models is considered

Theorem 2.1. *The finite-dimensional distributions of CSN AR models is CSN distribution. Also CSN AR models support the Markov property.*

Proof. Since for the observed data $\mathbf{X} = (X_1, \dots, X_n)$, we have $A\mathbf{X} \stackrel{d}{=} \mathbf{y}$ with

$$A = \begin{pmatrix} 1 & & & & & \\ -\phi & 1 & & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & -\phi & 1 \end{pmatrix}, \tag{2.2}$$

the density for \mathbf{X} is derived from closure property of the CSN distribution under linear transformations as:

$$\begin{aligned} p(\mathbf{x}) &= CSN_{n,n}(0, 2Q^{-1}, A, \mathbf{0}, 2I_n) \\ &= 2^n \phi_n(\mathbf{x}; 0, 2Q^{-1}) \Phi_n(A\mathbf{x}; \mathbf{0}, 2I_n). \end{aligned} \tag{2.3}$$

where

$$Q = A'A = \begin{pmatrix} 1 + \phi^2 & -\phi & & & & \\ -\phi & 1 + \phi^2 & -\phi & & & \\ & & \ddots & \ddots & \ddots & \\ & & & & -\phi & 1 + \phi^2 & -\phi \\ & & & & -\phi & 1 & \end{pmatrix}. \tag{2.4}$$

For considering the Markov property, we calculate $\pi(x_t|x_1, \dots, x_{t-1})$. Let $\mathbf{x}_t = (x_1, \dots, x_{t-1}, x_t)'$ and A_t and Q_t are $t \times t$ matrices with similar structure with matrices A and Q in (2.2) and (2.4), respectively. Now, we have

$$\begin{aligned} p(x_t|x_1, \dots, x_{t-1}) &= \frac{p(x_1, \dots, x_{t-1}, x_t)}{p(x_1, \dots, x_{t-1})} \\ &= 2 \frac{\phi_t(\mathbf{x}_t; 0, 2Q_t^{-1}) \Phi_t(A_t \mathbf{x}_t; \mathbf{0}, 2I_t)}{\phi_{t-1}(\mathbf{x}_{t-1}; 0, 2Q_{t-1}^{-1}) \Phi_{t-1}(A_{t-1} \mathbf{x}_{t-1}; \mathbf{0}, 2I_{t-1})} \\ &= 2\phi(x_t; \phi x_{t-1}, 2) \Phi((x_t - \phi x_{t-1}); 0, 2) \\ &= CSN_{1,1}(\phi x_{t-1}, 2, , 0, 2) = p(x_t|x_{t-1}) \end{aligned}$$

□

This is one of the main results. It simply says that similar to the Gaussian AR models, CSN AR models is also support the Markov property. Also when skewness parameter is zero, the density (2.3) reduces to the multivariate normal distribution.

Alternatively, using the representation of Dominguez-Molina et al. (2003) for CSN families, a stochastic representation of \mathbf{X} is

$$\mathbf{X} \stackrel{d}{=} \frac{1}{\sqrt{1+\pi^2}} A^{-1} |\mathbf{U}| + \frac{1}{\sqrt{1+\pi^2}} \mathbf{V}, \quad (2.5)$$

where $\mathbf{U} \sim N_n(0, I_n)$ and $\mathbf{V} \sim N_n(0, Q^{-1})$ are independent. From a computational point of view, this representation is useful because it implies that a random vector distributed according to (2.3) can be generated using two independent normal random vectors. The mean vector and covariance matrix of \mathbf{X} is

$$\begin{aligned} E(\mathbf{X}) &= \frac{1}{\sqrt{1+\pi^2}} \sqrt{\frac{2}{\pi}} A^{-1}, \\ Cov(\mathbf{X}) &= \frac{2}{\pi} \left(1 - \frac{2}{1+\pi^2}\right) Q^{-1} \end{aligned}$$

Hence, the zero elements of inverse of covariance matrix \mathbf{X} are the same of those of the precision matrix Q and so the precision matrix \mathbf{X} is quite sparse. This is a nice and useful result.

3 Conclusions

In this paper, the CSN AR model was introduced that enjoys some of the useful properties of Gaussian AR models like Markov property. Finally, we can similarly develop the CSN AR models to that of higher order.

References

- Allard, D., Naveau, P. (2007). A new spatial skew-normal random field model, *Communications in Statistics*, **36**, 1821-1834.
- Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171-178.
- Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Association*, **61**, 579-602.
- Capitanio, A., Azzalini, A., Stanghellini, E. (2003). Graphical models for skew-normal variates, *Scandinavian Journal of Statistics*, **30**, 129-144.

- Dominguez-Molina, J., Gonzalez-Farias, G., Gupta, A.K. (2003). The multivariate closed skew-normal distribution, *Technical Report* 03-12. Department of Mathematics and Statistics, Bowling Green State University.
- Gonzalez-Farias, G., Dominguez-Molina, J., Gupta, A. (2004). The closed skew-normal distribution. In: Genton, M., ed. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman Hall/CRC, 25-42.
- Karimi, O., Mohammadzadeh, M. (2010). Bayesian spatial regression models with closed skew normal correlated errors and missing observations, *Statistical Papers*, **53**, 205-218
- Pourahmadi, M. (2007). Skew-normal ARMA models with nonlinear heteroscedastic predictors, *Communications in Statistics*, **36**, 1803-1819.
- Prates, M.O., Dey, D.K., Lachos, V.H. (2012). Generalized Skew-Normal/Independent Fields with Applications.



Test of independence for some well-known bivariate distributions

M. Zargar¹, H. Jabbari and M. Amini

Department of Statistics, Ordered and Spatial Data Center of Excellence, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract: In this paper, we study the dependence structure of some bivariate distribution functions based on dependence measures of [Kochar and Gupta \(1987\)](#) and then compare these measures with Spearman's rho and Kendall's tau. Moreover, the empirical power of a class of distribution-free tests is computed based on exact and asymptotic distribution of U -statistics. Our results are obtained from simulation work in some continuous bivariate distributions. Also, we apply examples to illustrate the results.

Keywords: Copula functions, Celebioglu-Cuadras copula, Gumbel-Barnett distribution, Gumbel's bivariate distribution, Negative quadrant dependence, U-Statistics.

Mathematics Subject Classification (2010): 62G10, 62G05, 62H20.

1 Introduction and preliminaries

Dependence structure of bivariate distributions plays a useful role in statistic and applied probability. Also, for many purposes to the knowledge of the nature of dependence, it is also important to test independence against quadrant dependence. [Kochar and Gupta \(1987\)](#) introduced a class of distribution-free tests for testing independence against quadrant dependence and evaluated the empirical power of this class of tests in [Block and Basu's \(1974\)](#) family of distributions. [Amini et al. \(2010, 2011\)](#) evaluated the empirical power of this class of tests in bivariate FGM distribution and compared their results with some competitor tests.

Definition 1.1. ([Lehmann \(1966\)](#)). Let X and Y be two random variables. X and Y are NQD if $P(X \leq x, Y \leq y) \leq P(X \leq x)P(Y \leq y)$, for all (x, y) in \mathbf{R}^2 .

In this paper, we compute Kochar and Gupta's (1987) dependence measures for some bivariate distributions and compare them with Spearman's rho and Kendall's tau. Also, we test the independence hypothesis $H_0 : H(x, y) = F(x)G(y)$ against $H_1 : H(x, y) < F(x)G(y)$, for all $x, y \in \mathbf{R}$, where H is distribution function of a absolutely continuous random vector (X, Y) with marginals F and G .

¹M. Zargar : zargar-m@yahoo.com

2 Dependence measures

In this section, we use the copula functions of Gumbel’s bivariate exponential distribution, Gumbel-Barnett distribution and Celebioğlu-Cuadras distributions family and then, compute the dependence measures for these copula functions.

Kochar and Gupta (1987) introduced a well-known dependence measure for testing independence hypothesis against QD as follows. Let k be a fixed integer and consider $d_k(x, y) = H^k(x, y) - F^k(x)G^k(y)$, for all $x, y \in \mathbf{R}$.

For all $k \geq 1$, if X and Y are independent, then $d_k(x, y) = 0$. If (X, Y) is NQD , then $d_k(x, y) < 0$. The following measure of deviation between independence and QD is considered as

$$D_k = \int_{\mathbf{R}^2} d_k(x, y)dH(x, y) = D_{1k} - D_{2k}, \quad \forall k \geq 1, \tag{2.1}$$

where $D_{1k} = \int_{\mathbf{R}^2} H^k(x, y)dH(x, y)$ and $D_{2k} = \int_{\mathbf{R}^2} F^k(x)G^k(y)dH(x, y)$.

Due to good properties of copula functions, we consider dependence structure of these distributions via its copula functions. The Spearman’s ρ_s and Kendall’s τ are formulated by copula function C as follows:

$$\rho_s = 12 \int_0^1 \int_0^1 C(u, v)dudv - 3 \quad \text{and} \quad \tau = 4 \int_0^1 \int_0^1 C(u, v)dC(u, v) - 1 \tag{2.2}$$

In the next subsections, we compute the mentioned dependence measures for these distributions.

2.1 Gumbel’s bivariate exponential distribution

The joint survival function of Gumbel’s bivariate exponential distribution is

$$\bar{H}_G(x, y) = e^{-x-y-\theta xy}, \quad x, y \geq 0, \quad 0 \leq \theta \leq 1, \tag{2.3}$$

where θ is its dependence parameter. Its copula function is

$$C_G(u, v) = u + v - 1 + (1 - u)(1 - v) \exp\{-\theta \ln(1 - u) \ln(1 - v)\}, \tag{2.4}$$

and its copula survival function is $\hat{C}_G(u, v) = uv \exp\{-\theta \ln(u) \ln(v)\}$.

Proposition 2.1. *Let (X, Y) be a random vector with Gumbel’s bivariate exponential survival function (2.3) whose underlying copula function is C_G in (2.4). Then for all $0 \leq \theta \leq 1$ and $k \geq 1$,*

$$D_{2k} = \frac{1}{\theta} \sum_{i=0}^k \sum_{j=0}^k \binom{k}{i} \binom{k}{j} ij(-1)^{i+j} \exp\left\{\frac{(i+1)(j+1)}{\theta}\right\} E_1\left(\frac{(i+1)(j+1)}{\theta}\right),$$

$$\begin{aligned}
D_{1k} &= \sum_{i=0}^k \sum_{j=0}^i \sum_{l=0}^j \binom{k}{i} \binom{i}{j} \binom{j}{l} \frac{(-1)^{i-j+l}}{\theta(k-i+1)^3(l+k-i+1)(k-j+1)} \\
&\times \left\{ [(i-k)(k-i+1)\theta + (i-j)l](k-j+1)(l+k-i+1) \exp \left\{ \frac{(l+k-i+1)(k-j+1)}{\theta(k-i+1)} \right\} \right. \\
&\times \left. E_1 \left(\frac{(l+k-i+1)(k-j+1)}{\theta(k-i+1)} \right) - \theta(k-i+1)[l(k-j+1) - (k-i+1)(l+k-i+1)] \right\},
\end{aligned}$$

where E_1 is the exponential integral function defined by $E_1(x) = \int_x^\infty t^{-1} e^{-t} dt$.

Remark 2.2. Let (X, Y) be a random vector with Gumbel's bivariate exponential distribution. If $\theta = 0$, then we have the independence of (X, Y) and $0 < \theta \leq 1$ leads to NQD of (X, Y) .

Proposition 2.3. Let (X, Y) be a random vector with Gumbel's bivariate exponential survival function (2.3) and corresponding copula distribution function C_G in (2.4). Then for all $0 \leq \theta \leq 1$,

$$\rho_s = \frac{12}{\theta} e^{\frac{4}{\theta}} E_1 \left(\frac{4}{\theta} \right) - 3, \quad \text{and} \quad \tau = -e^{\frac{2}{\theta}} E_1 \left(\frac{2}{\theta} \right).$$

2.2 Gumbel-Barnett family

According to Hutchinson and Lai (1990), the copula distribution function of Gumbel-Barnett family is

$$C_{GB}(u, v) = uv \exp\{-\theta \ln(u) \ln(v)\}, \quad 0 \leq \theta \leq 1, \quad (2.5)$$

where θ is its dependence parameter.

Proposition 2.4. Let (X, Y) be a random vector with Gumbel-Barnett distribution whose underlying copula function is C_{GB} in (2.5). Then for all $0 \leq \theta \leq 1$ and $k \geq 1$,

$$\begin{aligned}
D_{2k} &= \frac{1-k}{1+k} + \frac{k^2}{\theta} \exp \left\{ \frac{(k+1)^2}{\theta} \right\} E_1 \left(\frac{(k+1)^2}{\theta} \right), \\
D_{1k} &= \frac{1}{(k+1)^2} \left[1 - k \exp \left\{ \frac{k+1}{\theta} \right\} E_1 \left(\frac{k+1}{\theta} \right) \right].
\end{aligned}$$

Remark 2.5. In this family of copulas, $\theta = 0$ implies independence and $0 < \theta \leq 1$ leads to NQD.

Proposition 2.6. Let (X, Y) be a random vector with Gumbel-Barnett distribution whose underlying copula function is C_{GB} in (2.5). Then for all $0 \leq \theta \leq 1$,

$$\rho_s = \frac{12}{\theta} e^{\frac{4}{\theta}} E_1 \left(\frac{4}{\theta} \right) - 3 \quad \text{and} \quad \tau = -e^{\frac{2}{\theta}} E_1 \left(\frac{2}{\theta} \right).$$

Remark 2.7. The Spearman's ρ_s , Kendall's τ and D_1 in Gumbel-Barnett distribution and Gumbel's bivariate exponential distribution are the same.

2.3 Celebioğlu-Cuadras copula

Celebioğlu (1997) introduced the following copula via a practical way of generating comprehensive copulas and Cuadras (2009) carefully studied its properties.

$$C_{CC}(u, v) = uv \exp\{\theta(1 - u)(1 - v)\}, \quad -1 \leq \theta \leq 1. \tag{2.6}$$

Proposition 2.8. *Let (X, Y) be a random vector whose underlying copula function is C_{CC} in (2.6). Then for all $-1 \leq \theta \leq 1$ and $k \geq 1$,*

$$D_{2k} = \frac{1}{k+1} - k \sum_{i=0}^{\infty} \frac{\theta^i}{i!} \text{Beta}^2(k+1, i+1) \left[1 - \theta \frac{(i+1)(k+1)}{(k+i+2)^2} \right], \tag{2.7}$$

$$D_{1k} = \frac{1}{k+1} - k \sum_{i=0}^{\infty} \frac{(\theta(k+1))^i}{i!} \text{Beta}^2(k+1, i+1) \tag{2.8}$$

$$\times \left[1 - 2\theta \frac{(i+1)(k+1)}{(k+i+2)^2} + \theta^2 \left(\frac{(k+1)(i+1)}{(k+i+2)(k+i+3)} \right)^2 \right], \tag{2.9}$$

where $\text{Beta}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Proposition 2.9. *Let (X, Y) be a random vector whose underlying copula function is C_{CC} in (2.6). Then for all $-1 \leq \theta \leq 1$,*

$$\rho_s = 12 \sum_{i=1}^{\infty} \frac{\theta^i}{i!} \text{Beta}^2(2, i+1) \text{ and } \tau = 1 - 4 \sum_{i=0}^{\infty} \frac{(2\theta)^i}{i!} \text{Beta}^2(2, i+1) \left[1 - 4\theta \frac{i+1}{i+3} + 2\theta^2 \left(\frac{i+1}{(i+3)(i+4)} \right)^2 \right].$$

2.4 Relations between dependence measures

The relations between dependence measures ρ_s, τ and $D_k, k = 1, 2, 3, 4$ have been considered for three copulas for *NQD* case in Table 1. The results show that for all copulas, the dependence measures are decreasing with respect to $|\theta|$ and we have $-1 < \rho_s < \tau < D_1 < D_2 < D_3 < D_4 < 0$. Comparing the dependence measures, we see that Gumbel’s bivariate exponential copula and Celebioğlu-Cuadras copula have the highest and lowest dependence, respectively. Also according to Remark 2.7, ρ_s, τ and D_1 have the same ranges for Gumbel’s bivariate exponential copula and Gumbel-Barnett copula.

3 Independence test against negative quadrant dependence

Let $H_0 : \theta = 0$ and $H_1 : \theta \geq \eta$ for Gumbel’s bivariate exponential copula and Gumbel-Barnett copula and let $H_0 : \theta = 0$ and $H_1 : \theta \leq -\eta$ for Celebioğlu-Cuadras copula, where $\eta > 0$. H_0 is independence

Table 1: Acceptable range of the dependence measures for underlying copulas.

dependence measures	Copulas		
	Gumbel's bivariate exponential copula	Gumbel-Barnett copula	Celebioğlu-Cuadras copula
ρ_s	(-0.5238, 0)	(-0.5238, 0)	(-0.2961, 0)
τ	(-0.3613, 0)	(-0.3613, 0)	(-0.1982, 0)
D_1	(-0.0466, 0)	(-0.0466, 0)	(-0.0248, 0)
D_2	(-0.0197, 0)	(-0.0172, 0)	(-0.0116, 0)
D_3	(-0.0087, 0)	(-0.0072, 0)	(-0.0058, 0)
D_4	(-0.0041, 0)	(-0.0035, 0)	(-0.0031, 0)

hypothesis and H_1 implies that X and Y are NQD . For testing H_0 against H_1 , [Kochar and Gupta \(1987\)](#) introduced the measure of deviation D_k in (2.1). Suppose that

$$\varphi_{k+1} \{(X_1, Y_1), \dots, (X_{k+1}, Y_{k+1})\} = \begin{cases} 1, & \{\max(X_1, \dots, X_{k+1}), \max(Y_1, \dots, Y_{k+1})\} \\ & \text{belongs to the same pair of } (X, Y) \\ 0, & \text{otherwise.} \end{cases} \quad (3.1)$$

where (X_i, Y_i) , for $i = 1, \dots, n$ be a random sample of (X, Y) with common distribution function H . For the independence test, we use U-statistic estimator of D_{1k} , that is

$$U_n(k+1) = \frac{1}{\binom{n}{k+1}} \sum \varphi_{k+1} \{(X_{i_1}, Y_{i_1}), \dots, (X_{i_{k+1}}, Y_{i_{k+1}})\}, \quad (3.2)$$

where the sum is over all combinations of $(k+1)$ integers (i_1, \dots, i_{k+1}) chosen out of $(1, \dots, n)$.

Remark 3.1. *The small values of $U_n(k+1)$ are significant for testing H_0 against H_1 .*

[Kochar and Gupta \(1987\)](#) obtained the exact distribution of $U_n(k+1)$. Also, we use Theorem A. of [Serfling \(1980\)](#) and obtain asymptotic distribution of $U_n(k+1)$.

3.1 Simulation study and applied examples

In this section, we compute the empirical powers of the test for H_0 against H_1 ($\eta = 0.2$) based on test statistic $U_n(k+1)$ for $k = 1, 2, 3, 4$ via the simulation results at 5% level of significance. Notice that for $n = 6, 8, 15, 20, (50)$, the critical points are calculated based on the exact (the asymptotic) distribution of $U_n(k+1)$. The following results are observed for all copulas (see Tables 2-4):

- For all n, k , the empirical sizes are acceptable and close to 0.05. Also, The powers are acceptable when n and $|\theta|$ are large. In addition, for all n , the powers increase when $|\theta|$ increases and for all $|\theta|$, the powers increase when n increases.
- The maximum and minimum powers of the independence tests are achieved for Gumbel's bivariate exponential copula and Celebioğlu-Cuadras copula, respectively.

Table 2: Critical points (CP), empirical sizes ($\theta = 0$) and empirical powers for $H_1 : \theta \geq 0.2$ and marginals $U(0, 1)$ in Gumbel's bivariate exponential copula.

n	U Statistics	CP for NQD	θ					
			0	0.2	0.5	0.7	0.9	1
6	$U_n(2)$	0.2	0.0304	0.0459	0.0827	0.106	0.1434	0.1665
	$U_n(3)$	0.05	0.0453	0.0668	0.1205	0.1475	0.1885	0.2161
	$U_n(4)$	0	0	0	0	0	0	0
	$U_n(5)$	0	0	0	0	0	0	0
8	$U_n(2)$	0.25	0.0305	0.056	0.1152	0.1641	0.2242	0.2564
	$U_n(3)$	0.0714	0.040	0.0697	0.139	0.19	0.256	0.2837
	$U_n(4)$	0.0143	0.036	0.0607	0.1189	0.1616	0.2154	0.2379
	$U_n(5)$	0	0	0	0	0	0	0
15	$U_n(2)$	0.3428	0.0468	0.1093	0.2678	0.4051	0.5480	0.6143
	$U_n(3)$	0.1516	0.0516	0.1190	0.2870	0.4267	0.5562	0.6155
	$U_n(4)$	0.0659	0.0519	0.1161	0.2718	0.3968	0.5141	0.5709
	$U_n(5)$	0.0259	0.0526	0.1152	0.2624	0.3792	0.4912	0.5440
20	$U_n(2)$	0.3684	0.0497	0.1365	0.3560	0.5334	0.6894	0.7613
	$U_n(3)$	0.1789	0.0497	0.1444	0.3782	0.5513	0.6956	0.7628
	$U_n(4)$	0.0910	0.0502	0.1422	0.3619	0.5226	0.6568	0.7236
	$U_n(5)$	0.0450	0.0495	0.1404	0.3429	0.4909	0.6172	0.6791
50	$U_n(2)$	0.4196	0.0552	0.2364	0.7024	0.8997	0.9757	0.9914
	$U_n(3)$	0.2403	0.0567	0.2717	0.7584	0.9255	0.9812	0.9921
	$U_n(4)$	0.1503	0.0505	0.2577	0.7380	0.9088	0.9745	0.9865
	$U_n(5)$	0.0966	0.046	0.2317	0.6928	0.8750	0.9534	0.9734

Table 3: Critical points (CP), empirical sizes ($\theta = 0$) and empirical powers for $H_1 : \theta \geq 0.2$ and marginals $U(0, 1)$ in Gumbel-Barnett copula.

n	U Statistics	CP for NQD	θ					
			0	0.2	0.5	0.7	0.9	1
6	$U_n(2)$	0.2	0.0297	0.0431	0.0808	0.1116	0.1470	0.1658
	$U_n(3)$	0.05	0.0464	0.0662	0.1071	0.1414	0.1783	0.2016
	$U_n(4)$	0	0	0	0	0	0	0
	$U_n(5)$	0	0	0	0	0	0	0
8	$U_n(2)$	0.25	0.031	0.0582	0.1164	0.164	0.2265	0.2519
	$U_n(3)$	0.0714	0.0398	0.0714	0.1238	0.1656	0.2318	0.2571
	$U_n(4)$	0.0143	0.0355	0.0594	0.1018	0.1360	0.1893	0.2120
	$U_n(5)$	0	0	0	0	0	0	0
15	$U_n(2)$	0.3428	0.0464	0.1102	0.2624	0.4036	0.5496	0.6214
	$U_n(3)$	0.1516	0.0481	0.0998	0.2187	0.3353	0.4570	0.527
	$U_n(4)$	0.0659	0.0484	0.092	0.1895	0.2841	0.3849	0.4496
	$U_n(5)$	0.0259	0.0494	0.0889	0.1767	0.2615	0.3574	0.4155
20	$U_n(2)$	0.3684	0.0484	0.1339	0.3562	0.529	0.6872	0.7634
	$U_n(3)$	0.1789	0.0480	0.1138	0.2764	0.4161	0.5725	0.6573
	$U_n(4)$	0.0910	0.0494	0.1024	0.2313	0.3454	0.4835	0.5653
	$U_n(5)$	0.0450	0.0496	0.0949	0.2071	0.3070	0.4289	0.5041
50	$U_n(2)$	0.4196	0.0504	0.2467	0.7055	0.9017	0.9761	0.9902
	$U_n(3)$	0.2403	0.0511	0.1897	0.5591	0.7858	0.9280	0.9668
	$U_n(4)$	0.1503	0.0487	0.1427	0.4231	0.6475	0.8394	0.9172
	$U_n(5)$	0.0966	0.0438	0.1137	0.3268	0.5291	0.7425	0.8398

Table 4: Critical points (CP), empirical sizes ($\theta = 0$) and empirical powers for $H_1 : \theta \geq 0.2$ and marginals $U(0, 1)$ in Celebioğlu-Cuadras copula.

n	U Statistics	CP for NQD	θ					
			0	-0.2	-0.5	-0.7	-0.9	-1
6	$U_n(2)$	0.2	0.0288	0.0355	0.0494	0.0566	0.0690	0.0760
	$U_n(3)$	0.05	0.0453	0.0545	0.0758	0.0848	0.0990	0.1086
	$U_n(4)$	0	0	0	0	0	0	0
	$U_n(5)$	0	0	0	0	0	0	0
8	$U_n(2)$	0.25	0.0294	0.0407	0.0632	0.0795	0.0943	0.1070
	$U_n(3)$	0.0714	0.04	0.0534	0.0763	0.0971	0.1143	0.128
	$U_n(4)$	0.0143	0.0362	0.0458	0.0655	0.0827	0.0985	0.1098
	$U_n(5)$	0	0	0	0	0	0	0
15	$U_n(2)$	0.3428	0.0447	0.0758	0.1234	0.1685	0.2222	0.2558
	$U_n(3)$	0.1516	0.0486	0.0750	0.1211	0.1678	0.2199	0.2515
	$U_n(4)$	0.0659	0.0477	0.0698	0.1168	0.1578	0.2052	0.2311
	$U_n(5)$	0.0259	0.0478	0.0696	0.1133	0.1521	0.1974	0.2199
20	$U_n(2)$	0.3684	0.0474	0.0847	0.1582	0.2181	0.2902	0.3366
	$U_n(3)$	0.1789	0.0489	0.0806	0.1491	0.2108	0.2834	0.3276
	$U_n(4)$	0.0910	0.0506	0.0779	0.1378	0.1968	0.2625	0.3002
	$U_n(5)$	0.0450	0.0508	0.0757	0.1324	0.1858	0.2436	0.2752
50	$U_n(2)$	0.4196	0.0474	0.1174	0.2881	0.4215	0.5901	0.6737
	$U_n(3)$	0.2403	0.0487	0.1162	0.2745	0.4204	0.5877	0.6701
	$U_n(4)$	0.1503	0.0442	0.1005	0.2390	0.3690	0.5284	0.6085
	$U_n(5)$	0.0966	0.0382	0.0835	0.1972	0.3095	0.4565	0.5341

Also separately for each copula, we have the following results:

- A) In Gumbel’s bivariate exponential copula, for all θ and n , $U_n(3)$ is the best test (see Table 2).
- B) In Gumbel-Barnett copula and Celebioğlu-Cuadras copula, when $n = 6, 8$ (small samples), $U_n(3)$ is the best test and for $n = 15, 20, 50$ (medium and large samples), $U_n(2)$ is the best one (see Tables 3 and 4).

3.2 Examples

We use the uranium exploration data set that was analysed already by Cook and Johnson (1986) (see data sets in *R* software for more details). For goodness of fit test in two dimensions, Peacock (1983) and Fasano and Franceschini (1987) proposed a generalization of the classical Kolmogorov-Smirnov test. We use the first 100 cases from some selected pairs in the uranium exploration data. Table 5 includes the maximum likelihood estimator (MLE) of θ , negative log-likelihood ($-\ell$) of MLE (θ), goodness of fit statistic (D_{ks}) and associated p-value. At 0.05 level, Table 5 shows that Gumbel’s bivariate exponential, Gumbel-Barnett and Celebioğlu-Cuadras distributions with 0.233, 0.244 and -0.973 for $MLE(\theta)$ are acceptable for the pairs (K, Sc) , (U, Sc) and (Co, K) , respectively.

Now, we test the independence against the strict NQD . The results are:

- A) For pairs of (K, Sc) and (U, Sc) according to the considered data, we test $H_0 : \theta = 0$ vs. $H_1 : \theta \geq 0.2$ ($\eta = 0.2$) at level $\alpha = 0.05$. So, we compute the test statistics $U_n(k + 1)$ for $k = 1, 2, 3, 4$, using the marginal empirical distribution function. The results is summarized in Table 6. Table 6 shows that

Table 5: MLE of θ , negative log-likelihood, the goodness of fit statistic D_{ks} and associated p-value for selected pairs of variables in the uranium exploration data

Pair	Model	$MLE(\theta)$	$-\ell$	D_{ks}	p-value
(U, Sc)	Gumbel's bivariate exp.	0.120	-1.099	1.066	0.533
	Gumbel-Barnett	0.244	-1.512	0.991	0.655
	Celebioğlu-Cuadras	-0.443	-1.143	9.797	0.026
(Co, K)	Gumbel's bivariate exp.	0.276	-3.438	0.816	0.892
	Gumbel-Barnett	0.211	-1.196	0.899	0.783
	Celebioğlu-Cuadras	-0.973	-3.160	7.478	0.992
(K, Sc)	Gumbel's bivariate exp.	0.233	-2.704	0.777	0.916
	Gumbel-Barnett	0.131	-0.502	0.895	0.792
	Celebioğlu-Cuadras	-1	-4.284	9.702	0.034

Table 6: Values of test statistics, their critical values and associated p-values for selected pairs of variables in the uranium exploration data

Pair	Model		$U_n(2)$	$U_n(3)$	$U_n(4)$	$U_n(5)$
(K, Sc)	Gumbel's bivariate exp.	Test statistic	0.409	0.217	0.129	0.081
		Critical value	0.445	0.267	0.179	0.127
		p-value	0.003	0.002	0.002	0.004
(U, Sc)	Gumbel-Barnett	Test statistic	0.448	0.298	0.228	0.183
		Critical value	0.445	0.267	0.179	0.127
		p-value	0.058	0.187	0.304	0.355
(Co, K)	Celebioğlu-Cuadras	Test statistic	0.427	0.236	0.142	0.089
		Critical value	0.445	0.267	0.179	0.127
		p-value	0.014	0.007	0.006	0.006

- for pair (K, Sc) based on Kochar and Gupta's tests, we reject H_0 in favor of H_1 . Since $MLE(\theta) = 0.233$, the result of independence test is true.
- for pair (U, Sc) based on Kochar and Gupta's tests, we can't reject H_0 in favor of H_1 .

B) For pair of (Co, K) according to the considered data, we test $H_0 : \theta = 0$ vs. $H_1 : \theta \leq -0.7$ ($\eta = -0.7$) at level $\alpha = 0.05$. Table 6 shows that for pair (Co, K) based on Kochar and Gupta's tests (1987), we reject H_0 in favor of H_1 . So, Co and K are not independent. According to $MLE(\theta)$, we conclude that the results of independence tests are coincident with the truth.

Remark 3.2. According to Table 6, we conclude that Kochar and Gupta's tests are good for detecting independence against NQD in Gumbel's bivariate exponential and Celebioğlu-Cuadras distributions.

References

Amini, M., Jabbari, H., Mohtashami Borzadaran, G. R., Azadbakhsh, M. (2010). Power comparison of independence test for the Farlie-Gumbel-Morgenstern family, *Communications of the Korean Statistical Society* **17**(4), 493–505.

- Amini, M., Jabbari, H., Mohtashami Borzadaran, G. R. (2011). Aspects of dependence in generalized Farlie-Gumbel-Morgenstern distributions, *Communications in Statistics-Simulation and Computation* **40**, 1192–1205.
- Block, H. W., Basu, A. P. (1974). A continuous bivariate exponential extension, *Journal of the American Statistical Association* **69**, 1031–1037.
- Celebioglu, S. (1997). A way of generating comprehensive copulas, *Journal of the Institute of Science and Technology* **10** (1), 57–61.
- Cook, R. D., Johnson, M. E. (1986). Generalized Burr-Pareto-Logistic Distributions With Applications to a Uranium Exploration Data Set, *Technometrics* **28**(2), 123–131.
- Cuadras, C. M. (2009). Constructing copula functions with weighted geometric means, *Journal of Statistical Planning and Inference* **139** , 3766-3772.
- Fasano, G., Franceschini, A. (1987). A multidimensional version of the Kolmogorov-Smirnov test, *Mon. Not. R. astr. Soc.* **225**, 155-170.
- Hutchinson, T. P., Lai, C. D. (1990). Continuous bivariate distributions, Emphasising Applications. *Rumsby Scientific Publishing*, Adelaide.
- Kochar, S. G., Gupta, R. P. (1987). Competitors of Kendall-tau test for testing independence against positive quadrant dependence, *Biometrika* **74** (3), 664–666.
- Lehmann, E. L. (1966). Some concepts of dependence, *Annals of Mathematics and Statistics* **37**, 1137–1153.
- Peacock, J. A. (1983). Two-dimensional goodness-of-fit testing in astronomy, *Mon. Not. R. astr. Soc.* **202**, 615-627.
- Serfling, R. J. (1980). *Approximations Theorems of Mathematical Statistics*, John Wiley & Sons.



The Topp-Leone Distribution: Revisited

A. Pourdarvish, S.M.T.K. MirMostafaei, Kimia Naderi¹

Department of Statistics, University of Mazandaran, Babolsar, Iran

Abstract: deals with a two-parameter model, called the Topp-Leone distribution. Some of the main properties of this important model, such as the shape of the density, hazard rate function, mean residual lifetime and the moments have been reviewed. Maximum likelihood estimation of the parameters is also discussed. The paper concludes with a real data example.

Keywords Akaike information criterion, Mean residual lifetime, Moments, Reverse hazard rate function.

Mathematics Subject Classification (2010): 60E05; 62E99.

1 Introduction

In this paper, we consider the two-parameter Topp-Leone distribution. We study several interesting properties of this distribution. [Topp and Leone \(1955\)](#) proposed a family of univariate distributions by formulating the cumulative distribution function (cdf), $F(x)$, as

$$F(x; \nu, b) = \left(\frac{x}{b}\right)^\nu \left(2 - \frac{x}{b}\right)^\nu, \quad 0 \leq x \leq b. \quad (1.1)$$

where ν and b are the shape and scale parameters, respectively. [Nadarajah and Kotz \(2003\)](#) provided a motivation for this distribution based on its hazard rate function and then investigated some properties of the Topp-Leone distribution. Some reliability measures and asymptotic distribution of order statistics of this distribution have been discussed by [Ghitany \(2007\)](#) and [Ghitany et al. \(2007\)](#), respectively. [Genc \(2012\)](#) and [MirMostafaei \(2014\)](#) focused on the moments of order statistics from the Topp-Leone distribution.

In what follows, we review some distributional properties of the Topp-Leone distribution in Section 2. The moments are discussed in Section 3. Section 4 includes a discussion regarding the maximum likelihood estimation of the parameters. In the end, a real data example is provided to illustrate the usefulness of this model.

¹Speaker's name: naderi.kimia@yahoo.com

2 Distributional Properties

If the random variable X has a Topp-Leone distribution as defined in (1.1), then it has the probability density function (pdf)

$$f(x; \nu, b) = \frac{2\nu}{b} \left(1 - \frac{x}{b}\right) \left(\frac{x}{b}\right)^{\nu-1} \left(2 - \frac{x}{b}\right)^{\nu-1}, \quad 0 \leq x \leq b, \quad 0 < \nu < 1. \quad (2.1)$$

Under the restriction $0 < \nu < 1$, the pdf is a convex function with a unique mode 0 at $x = 0$ and it continuously decreases to 0 as x approaches to b from the left. The distribution is therefore reversed J-shaped throughout its support.

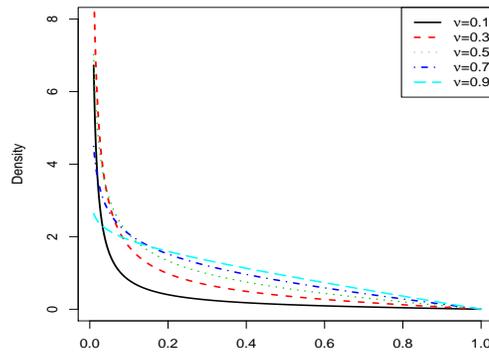


Figure 1: Plots of $f(x)$ in equation (2.1) for $\nu = 0.1, 0.5, 1, 2, 5$ and $b = 1$.

Now let us look at the hazard rate function of the Topp-Leone distribution. Let $y = 1 - \frac{x}{b}$, then the hazard rate function is given by

$$h(x) = \frac{2\nu}{b} \frac{y(1-y^2)^{\nu-1}}{1 - (1-y^2)^\nu}, \quad (2.2)$$

After some calculations, we observe that $h(x)$ attains a minimum at $x = x_0$, where $y_0 = 1 - \frac{x_0}{b}$ is the root of the equation.

Figure 2 illustrates the shape of equation (2.2) for a range of ν and when $b = 1$. It is evident that equation (2.2) has the 'bathtub' shape for the full range $\nu \in (0, 1)$.

Now we discuss the reverse hazard rate of the Topp-Leone distribution. The reverse hazard rate of any distribution function $F(x)$, can be defined as $a(x) = \frac{f(x)}{F(x)}$. Consequently, the reversed hazard

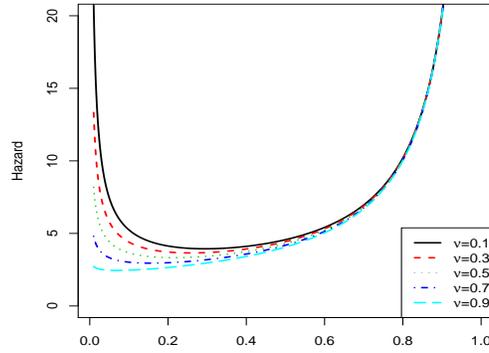


Figure 2: Plots of $h(x)$ in equation (2.2) for $\nu = 0.1, 0.5, 1, 2, 5$ and $b = 1$.

rate of this distribution, is given by

$$a(x) = \frac{2\nu(b - x)}{x(2b - x)}, \quad 0 < x < b \tag{2.3}$$

The reverse hazard rate has recently become quite popular in the statistical literature. It can be interpreted as follows. Suppose, the lifetime of a unit has reverse hazard rate $a(x)$, then $a(x)dx$ provides the probability of the lifetime failing in $(x - dx, x)$, when a unit is found failed at x . In general, the reverse hazard rate is useful in constructing the information matrix and in estimating the survival function for censored data, see Block et al. (1998). The mean residual life (MRL) is defined as

$$\mu(x) = E(X - x | X > x) = \frac{1}{\bar{F}(x)} \int_x^\infty \bar{F}(y) dy.$$

For a unit having already survived up to time x , $\mu(x)$ measures its expected remaining lifetime. Evidently $\mu(0) = E(X) = \mu$ is the mean of X .

Let $Be(a, b) = \int_0^1 t^{a-1}(1 - t)^{b-1} dt$ and $Be(a, b; c) = \int_0^c t^{a-1}(1 - t)^{b-1} dt$ be the complete and incomplete beta functions, respectively. Since for the Topp-Leone distribution, $\bar{F}(x) = 0$ for $x \geq b$, using the substitution $z = \frac{y}{2b}$ and the fact that $2Be(a, a; \frac{1}{2}) = Be(a, a)$, we have

$$\begin{aligned} \int_x^b \bar{F}(y) dy &= b - x - \int_x^b \left(\frac{y}{b}\right)^\nu \left(2 - \frac{y}{b}\right)^\nu dy = b - x - b4^{\nu+\frac{1}{2}} \int_{\frac{x}{2b}}^{\frac{1}{2}} z^\nu (1 - z)^\nu dy \\ &= b - x - b4^{\nu+\frac{1}{2}} \left\{ Be\left(1 + \nu, 1 + \nu; \frac{1}{2}\right) - Be\left(1 + \nu, 1 + \nu; \frac{x}{2b}\right) \right\} \end{aligned}$$

$$= b - x - b4^\nu \left\{ Be\left(1 + \nu, 1 + \nu\right) - 2Be\left(1 + \nu, 1 + \nu; \frac{x}{2b}\right) \right\}. \quad (2.4)$$

Therefore, we have

$$\mu(x) = \frac{b - x - b4^\nu \left\{ Be\left(1 + \nu, 1 + \nu\right) - 2Be\left(1 + \nu, 1 + \nu; \frac{x}{2b}\right) \right\}}{1 - \left(\frac{x}{b}\right)^\nu \left(2 - \frac{x}{b}\right)^\nu}. \quad (2.5)$$

Now consider the following theorem which has been stated and proved by [Ghitany et al. \(2007\)](#).

Theorem 2.1. *If $X \sim TL(\nu, b)$ and $\nu \in (0, 1)$, then X has bathtub hazard rate function (upside-down bathtub MRL).*

3 The moments

The moment generating function of $X \sim TL(\nu, b)$ is given by

$$\begin{aligned} M(t) &= E(e^{tX}) = \int_0^b e^{tx} f_{TL}(x) dx \\ &= \sum_{m=0}^{\infty} \frac{2\nu t^m}{b m!} \int_0^b x^m \left(1 - \frac{x}{b}\right) \left(\frac{x}{b}\right)^{\nu-1} \left(2 - \frac{x}{b}\right)^{\nu-1} dx \\ &= \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} \binom{\nu-1}{i} \frac{2\nu t^m}{b m!} \int_0^b x^m \left(1 - \frac{x}{b}\right)^{i+1} \left(\frac{x}{b}\right)^{\nu-1} dx \\ &= \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{i+1} \binom{\nu-1}{i} \binom{i+1}{j} (-1)^{i+1} \frac{2\nu t^m}{b^{\nu+j} m!} \int_0^b x^{m+\nu+j-1} dx \\ &= \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{i+1} \binom{\nu-1}{i} \binom{i+1}{j} (-1)^{i+1} \frac{2\nu t^m}{m!} \frac{b^{m-1}}{m + \nu + j}. \end{aligned} \quad (3.1)$$

It was used in the proof that $e^{tx} = \sum_{m=0}^{\infty} \frac{(tx)^m}{m!}$ and $(1-z)^a = \sum_{i=0}^{\infty} \binom{a}{i} (-z)^i$.

It follows that the n -th moment of X exists and it is given by ($n \in N$);

$$\begin{aligned} E(X^n) &= (-1)^{1+n} b^n \frac{(-1-n-\nu)_{1+n}}{(1+\nu)_{1+n}} - nb^n 2^{n+2\nu-1} Be(1+\nu, n+\nu) \\ &\quad - \frac{\nu b^n}{n+\nu} \sum_{k=1}^n (-1)^k (k-1) \frac{(-n-\nu)_{k-1}}{(\nu)_k}. \end{aligned} \quad (3.2)$$

Let X_1, \dots, X_n be a random sample of size n from the Topp-Leone distribution and let $X_{1:n} < \dots < X_{n:n}$ denote the order statistics obtained from this sample. The r -th order statistic, $X_{r:n}$, represents

the life length of a $(n - r + 1)$ -out of n system made up of n identical components with independent life lengths. The pdf of the r -th order statistic, $X_{r:n}$, denoted by $f_{r:n}(x)$, for $1 \leq r \leq n$ from the Topp-Leone distribution is

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} \left[\left(\frac{x}{b} \right) \left(2 - \frac{x}{b} \right) \right]^{\nu(r-1)} \left[1 - \left(\frac{x}{b} \right)^\nu \left(2 - \frac{x}{b} \right)^\nu \right]^{n-r} \frac{2\nu}{b} \left(1 - \frac{x}{b} \right) \left[\left(\frac{x}{b} \right) \left(2 - \frac{x}{b} \right) \right]^{\nu-1}.$$

The k -th moment of the r -th order statistic, $\mu_{r:n}^{(k)} = E(X_{r:n}^k)$, $k \in N$, $1 \leq r \leq n$, is as follows

$$\begin{aligned} \mu_{r:n}^{(k)} &= \frac{n!a^k}{(r-1)!(n-r)!} \sum_{j=0}^{n-r} \binom{n-r}{j} (-1)^j 2^{k+2a(r+j)} \\ &\quad \times [Be(k+a(r+j), a(r+j), 1/2) - 2Be(k+a(r+j)+1, a(r+j), 1/2)]. \end{aligned} \tag{3.3}$$

Also, [MirMostafaei \(2014\)](#) obtained the following relation

$$\mu_{r+1:n}^{(k+1)} = \left(1 + \frac{1}{k} \right) \mu_{r+1:n}^{(k)} + \left(\frac{k+1}{2ar} + 1 \right) \mu_{r:n}^{(k+1)} - \left(\frac{k+1}{ar} + \frac{1}{k} + 1 \right) \mu_{r:n}^{(k)}.$$

The joint pdf of any two order statistics $U = X_{r:n}$ and $V = X_{s:n}$ ($1 \leq r < s \leq n$) is given by

$$\begin{aligned} f_{r,s:n}(u,v) &= C_{r,s,n} 4\nu^2 \left[\frac{u}{b} \left(2 - \frac{u}{b} \right) \right]^{\nu i-1} \left[\frac{v}{b} \left(2 - \frac{v}{b} \right) \right]^{\nu-1} \\ &\quad \times \left[\left[\frac{v}{b} \left(2 - \frac{v}{b} \right) \right]^\nu - \left[\frac{u}{b} \left(2 - \frac{u}{b} \right) \right]^\nu \right]^{s-r-1} \left[1 - \left(\frac{v}{b} \left(2 - \frac{v}{b} \right) \right)^\nu \right]^{n-s} \left(1 - \frac{u}{b} \right) \left(1 - \frac{v}{b} \right). \end{aligned}$$

where $C_{r,s,n} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$.

Therefore, we define the product moment of the two order statistics as

$$\mu_{r,s:n}^{(k,l)} = E(X_{r:n}^k X_{s:n}^l).$$

$$\begin{aligned} \mu_{r,s:n}^{(k,l)} &= C_{r,s,n} \sum_{i=0}^{n-s} \sum_{j=0}^{s-r-1} \sum_{q=0}^{\infty} \binom{n-s}{i} \binom{s-r-1}{j} \binom{(r+j)\nu-1}{q} (-1)^{i+j+q} 2^{\ell+k+2(s+i)\nu-2} \\ &\quad \times \left[\frac{1}{\xi_{r,k,\nu}(j,q)} Be(\xi_{s,k,\ell,\nu}^*(i,q), \nu A_{i,j}, 1/2) \right. \\ &\quad - \frac{2[2\xi_{r,k,\nu}(j,q)+1]}{\xi_{r,k,\nu}(j,q)[\xi_{r,k,\nu}(j,q)+1]} Be(\xi_{s,k,\ell,\nu}^*(i,q)+1, \nu A_{i,j}, 1/2) \\ &\quad \left. + \frac{4}{\xi_{r,k,\nu}(j,q)+1} Be(\xi_{s,k,\ell,\nu}^*(i,q)+2, \nu A_{i,j}, 1/2) \right], \end{aligned}$$

where $\xi_{r,k,\nu}(j,q) = k + (j+r)\nu + q$ and $\xi_{s,k,\ell,\nu}^*(i,q) = \ell + k + q + (s+i)\nu$. If $a \in N$, then the third summation stops at $q = (j+r)\nu - 1$.

4 Maximum likelihood estimation

In this section, we briefly discuss the maximum likelihood estimators (MLE's) of the two-parameter TL distribution and discuss their asymptotic properties.

Let X_1, \dots, X_n be a random sample from TL, the log-likelihood function can be written as;

$$\ell(\nu, b) = n \ln\left(\frac{2\nu}{b}\right) + \sum_{i=1}^n \ln\left(1 - \frac{x_i}{b}\right) + (\nu - 1) \left[\sum_{i=1}^n \ln\left(\frac{x_i}{b}\right) + \sum_{i=1}^n \ln\left(2 - \frac{x_i}{b}\right) \right]. \quad (4.1)$$

Now, by differentiating the log-likelihood function with respect to the parameters and equating the results with zero, we have

$$\begin{aligned} \frac{\partial \ell}{\partial b} &= -n + \sum \frac{x_i}{b - x_i} + (\nu - 1) \left[-n + \sum \frac{x_i}{2b - x_i} \right] = 0. \\ \frac{\partial \ell}{\partial \nu} &= \frac{n}{\nu} + \sum_{i=1}^n \ln\left(\frac{x_i}{b}\right) + \sum_{i=1}^n \ln\left(2 - \frac{x_i}{b}\right) = 0. \end{aligned} \quad (4.2)$$

From the above likelihood equations, we can see that for a given b , the MLE of ν is

$$\widehat{\nu}(b) = -\frac{n}{\sum_{i=1}^n \ln\left(\frac{x_i}{b}\right) + \sum_{i=1}^n \ln\left(2 - \frac{x_i}{b}\right)}. \quad (4.2)$$

Substituting (4.2) in (4.2), we can obtain the MLE of b . Once we get the MLE of b , the MLE of ν can be obtained as $\widehat{\nu}(\widehat{b})$ from (4.2).

5 Application

This example deals with the monthly water capacity data from the Shasta reservoir in California, USA and were taken for the month of February from 1995 to 2010, see <http://cdec.water.ca.gov/reservoir-map.html>. The maximum capacity of the reservoir is 4552000 AF and the data were transformed to the interval $[0,1]$. These data are presented in Table 1. We fitted the following 3 distributions to the above data: the standard Topp-Leone distribution; the Marshall-Olkin uniform distribution whose pdf is

$$g_1(x, \nu) = \frac{\nu}{[\nu + (1 - \nu)x]^2}, \quad x \in (0, 1),$$

the Power distribution whose pdf is

$$g_2(x, \nu) = \nu x^{\nu-1}, \quad x \in (0, 1).$$

Table 2 lists the parameter estimates, their standard errors, the negative log-likelihood values (-LL), the values of the Akaike information criterion (AIC), the values of the Bayesian information criterion (BIC), the Kolmogorov-Smirnov (K-S) test statistics and their corresponding p -values.

Table 1: Monthly capacity for August and proportion of capacity for Shasta reservoir.

Year	Capacity	proportion of total capacity	Year	Capacity	proportion of total capacity
1995	1542838	0.338936	2003	3584283	0.787408
1996	3694201	0.811556	2004	3868600	0.849868
1997	3574861	0.785339	2005	3168056	0.695970
1998	3567220	0.783660	2006	3834224	0.842316
1999	3712733	0.815627	2007	3772193	0.828689
2000	3857423	0.847413	2008	2641041	0.580194
2001	3495969	0.768007	2009	1960458	0.430681
2002	3839544	0.843485	2010	3380147	0.742563

We also plotted the empirical cdf plots of the fitted models. From Table 2 and Figure 3, we observe that the Topp-Leone distribution has the smallest -LL value, the smallest AIC, BIC K-S test statistic values and the largest p -value and therefore provides the best fit among the distribution considered here for the data.

Table 2: The MLE's and the measures for the fitted models.

Model	MLE	Error	-LL	AIC	BIC	K-S	p -value
Topp-Leone	13.2748	3.3187	-13.5925	-25.1850	-24.4124	0.2611	0.1885
Marshall-Olkin Uniform	3.4371	1.2514	-5.4249	-8.5640	-8.0771	0.3778	0.0146
Power	3.4908	0.8727	-8.5856	-15.1713	-14.3987	0.4333	0.0030

References

- Block, H., Savits, T.H. and Singh, H. (1998). The reversed hazard rate function, *Probability in the Engineering and Informational Sciences*, **12**, 69–90.
- Genç A.İ. (2012). Moments of Order Statistics of Topp-Leone Distribution, *Statistical Papers*, **53**, 117–131.
- Ghitany, M.E. (2007). Asymptotic distribution of order statistics from the Topp-Leone Distribution, *Journal of Applied Mathematics*, **20**, 371–376.

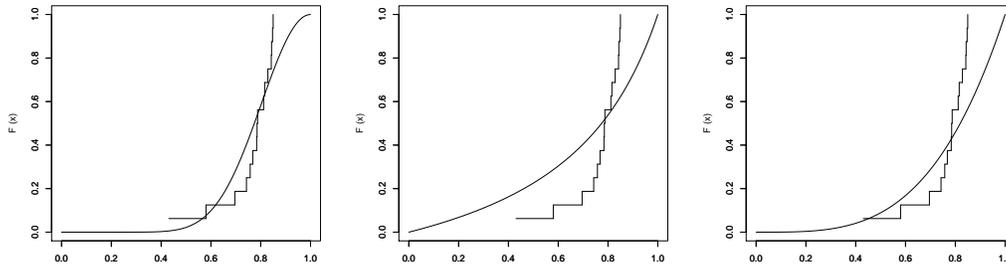


Figure 3: Empirical cdf plots of the fitted Topp-Leone (left), Marshall-Olkin uniform (center) and power (right) models.

Ghitany, M.E., Kotz, S. and Xie, M. (2005). On Some Reliability Measures and their Stochastic Orderings for the Topp-Leone Distribution, *Journal of Applied Statistics*, **32**, 715–722.

MirMostafaei, S.M.T.K. (2014). On the Moments of Order Statistics Coming from the Topp-Leone Distribution, *Statistics and Probability Letters*, **95**, 85–91.

Nadarajah, S. and Kotz, S. (2003). Moments of Some J-shaped Distributions, *Journal of Applied Statistics*, **30**, 311–317.

Topp, C.W. and Leone, F.C. (1955). A Family of J-shaped Frequency Functions, *Journal of the American Statistical Association*, **50**, 209–219.