In The Name of God

Proceedings

(Invited and refereed papers)

The 10^{th} Iranian Statistical Conference

August 3–5, 2010
University of Tabriz
Tabriz, Iran

Proceedings of The 10th Iranian Statistical Conference
Preparation: LaTeX, Javad Mehri-Tekmeh (Ph.D.)
Publication: University of Tabriz
Pub. Date: August 2010

Perface

The Proceedings of the tenth Iranian Statistical Conference, published in two volumes, is the product of sustained endeavors made by members of the Scientific and Organizing Committees, scientific advisers, several students and personnel of the Faculty of Mathematical sciences on one hand, and the authors of papers on the other hand. Manuscripts appearing in this volume are invited and accepted contributed papers which are written in English, arranged in alphabetic order of their first author. It should be noted that these papers are chosen by referees and the Scientific Committee from 262 out of 600 papers whose articles had been accepted for oral presentation. Invited papers have not been refereed. However, due to time limitations, no attempts have been made to have the papers revised by their authors or the Scientific Committee, except for some minor editing corrections. Therefore, the authors of papers are responsible for the content of these manuscripts. Further, for the same reasons, some contributed papers that have not reached the conference office on time or have not fulfilled the conference requirements are not appeared in the Proceedings.

The Conference Organizers

Organizing Committee

1. Aghajani, K.
2. Alavi, S.M.T. (University president)
3. Bevrani, H.
4. Emamalipour, H.
5. Ghasemzadeh, H.R.
6. Jabbari Khamneh, H.
7. Najafi, S.
8. Navid, H.
9. Novin, A.R.
10. Rostami, A.

Executive Committee

1. Adeli Shahir, M.
2. Ahmadi Shali, J.
3. Badamchizadeh, A.
4. Bevrani, H.
5. Emamalipour, H.
6. Haghighian, J.
7. Heydari, A.A.
8. Hosseini, S.S.
9. Imani, R.
10. Jabbari Khamneh, H.(Chairman)
11. Mehri Tekmeh, J.
12. Mousavi, H.
13. Parsian, A.
14. Shahri Heravi, M.

Scientific Committee

1. Bevrani, H.(Chairman)	University of Tabriz
2. Heydari, A.A.	University of Tabriz
3. Homei, H.	University of Tabriz
4. Imani, R.	University of Tabriz
5. Jabbari Khamneh, H.	University of Tabriz
6. Mehrabi, Y.	Shahid Beheshti University of Medical Sciences
7. Mohammadzadeh, M.	Tarbiat Modares University
8. Nematollahi, N.	Allameh Tabatabaei University
9. Parsian, A.	University of Tehran

CONTENTS

Ordering Comparison of Zero-truncated Poisson Random Variables with their Mixtures

S. Abbasi, M. Aghababaei Jazi and M. H. Alamatsaz

Naghshe Jahan, higher education institute, Isfahan, Iran

Faculty of Mathematics and Statistics, University of Sistan and Balochestan,

Department of Statistics, University of Isfahan

Comparison of random variable (r.v.'s) measuring certain characteristics in various types of stochastic orderings are of interest in diverse areas. In particular, several authors have been recently concerned with the comparison of Poisson, binomial, negative binomial and logarithmic series distributions with their respective mixtures. Incidentally, these distributions are among the four well-known distributions of the family of generalized power series (GPS) distributions. The objective of this article is to compare a zero- truncated Poisson distribution having a fixed parameter with a zero- truncated Poisson mixture having a random parameter, which has an arbitrary continuous (or discrete) probability distribution. Comparisons are made with respect to various types of orderings, such as, the simple stochastic, likelihood ratio, (reversed) hazard rate, uniformly more variable and expectation orderings. The special case when the means of the two distributions are the same is also discussed.

Keywords: Generalized power series distributions; Simple stochastic ordering; Expectation ordering; Likelihood ratio ordering; (reversed) hazard rate ordering; Uniformly more variable ordering; Convex ordering; Mean residual life ordering; (Factorial) Moments ordering; Moment generating function ordering;

1. Introduction

In the marketing literature, modeling the customer's response behavior to price and promotion effects is a major area of research. In this field, it is widely accepted that the consumer's quantity decision (i.e., how many units to buy within a product category) at a given purchase incidence can be modeled as a stochastic variable, following a zero-truncated Poisson distribution (see, e.g., Johnson and Kotz (1969) and Campo et al (2003)). Typically, the Poisson parameter in these models is estimated as a function of customer-specific variables (e.g., loyalty) and marketing variables e.g., price or promotion.

Hence, mixtures of distributions used in building probability models are quite frequently in biological sciences. For instance, in order to study certain characteristics in natural populations of fish, a random sample might be taken and the characteristic measured for each member of the sample; since the characteristic varies with the age of the fish, the distribution of the characteristic in the total population will be a mixture of the distribution at different ages. In order to analyse the qualitative character of inheritance, a geneticist might observe a phenotypic value that has a mixture distribution because each genotype might produce phenotypic values over an interval. For a comprehensive account on mixture of distributions as they occur in diverse fields we refer to Sapatinas, T (1994).
Now, consider that the probability density function (pdf) of some random variable X of interest is a mixture of pdf' s of the r.v.'s Y_θ, $\theta \in \Omega$,and it is of interest to compare r.v.'s X and Y_{θ_0} for a fixed $\theta_0 \in \Omega$. In particular, it may be of interest to compare r.v.'s X and Y_μ; where μ is the mean of the mixing distribution. Shaked (1980) made such comparisons when the pdf' s of r.v.'s Y_θ, $\theta \in \Omega$, belong to the general exponential family. He showed that the two resulting pdf's must cross each other exactly twice in a prescribed manner. Misra et al. (2003) considered Poisson and binomial distributions and compared them with their corresponding mixtures, with respect to various stochastic orderings. Motivated by Misra et al's work, Alamatsaz and Abbasi (2008) compared negative binomial distributions with their corresponding mixtures and obtained similar results. Recently, Aghababaei Jazi and Alamatsaz (2009) have studied the stochastic ordering comparison of another known member of generalized power series distributions, i.e., the logarithmic series distribution, with its mixture. Motivated by their work, in this paper, we consider a zero-truncated Poisson distribution and compare it with its mixture with respect to various stochastic orderings and obtain similar results. Precisely, let X be a r.v. having zero-truncated Poisson distribution with a fixed parameter $\lambda \in (0, \infty)$, i.e.,

$$P(X = k) = \begin{cases} \frac{\lambda^k}{k!(e^\lambda - 1)} & k = 1, 2, ... \\ 0 & o.w. \end{cases} \tag{1}$$

Let Y be a zero-truncated Poisson r.v. with a variable parameter θ, where θ is a non-degenerate r.v. having a probability density function $g(\theta)$, $\theta > 0$. (Although we have assumed θ to be a continuous r.v., all results obtained here would also hold when θ is a discrete r.v.). Thus, we have

$$P(Y = k) = \begin{cases} \int_0^\infty \frac{\theta^k}{k!(e^\theta - 1)}\, g(\theta)\, d(\theta), & k = 1, 2, ... \\ 0, & o.w. \end{cases} \tag{2}$$

Our purpose is to compare r.v.'s X and Y, given by (1) and (2), respectively, with respect to various types of stochastic orderings reviewed in section 2.

In section 3, we derive conditions which ensure that a mixed zero-truncated Poisson r.v., with an arbitrary mixing distribution, is larger than the corresponding zero-truncated Poisson r.v. with a fixed parameter in different senses. As a special case, the two distributions are compared when they have equal means.

2. Stochastic Orderings

Here, a brief account of some known stochastic orderings, used in this paper, is provided and their properties are stated. For more details, we refer to, e.g., Muller and Stoyan (2002) and Shaked and Shanthikumar (2007).
Let X and Y be two nonnegative discrete (or continuous) r.v.'s with respective distribution functions $F(t) = \Pr(X \leq t)$ and $G(t) = \Pr(Y \leq t)$, probability density functions $f(t) = \Pr(X = t)$ and $g(t) = \Pr(Y = t)$, survival functions $\bar{F}(t)$ and $\bar{G}(t)$, hazard (failure) rate functions $r_X(t) = f(t)/\Pr(X \geq t)$ and $r_Y(t) = g(t)/\Pr(Y \geq t)$ and reversed hazard rate functions $\bar{r}_X(t) = f(t)/\Pr(X \leq t)$ and $\bar{r}_Y(t) = g(t)/\Pr(Y \leq t)$.

2.1. Simple stochastic ordering: X is said to be smaller than Y in simple stochastic ordering, denoted by $X \leq_{st} Y$, if $G(t) \leq F(t)$ for all $t \geq 0$, or, equivalently, if $E(\phi(X)) \leq E(\phi(Y))$ for all non-decreasing functions ϕ for which the expectations exist. Thus, in particular, $X \leq_{st} Y$ implies that $E(X) \leq E(Y)$.

2.2. Expectation ordering: X is said to be smaller than Y in expectation ordering, denoted by $X \leq_E Y$, if $E(X) \leq E(Y)$, where expectations are assumed to exist. Thus, by 2.1, simple stochastic ordering implies expectation ordering but the converse may not be true.

2.3. Likelihood ratio ordering: X is said to be smaller than Y in likelihood ratio ordering, denoted by $X \leq_{lr} Y$, if $f(u)\, g(v) \geq f(v)\, g(u)$, for all $u \leq v$, which is equivalent to $h(t) = f(t)/g(t)$ being non-increasing. Also $X \leq_{lr} Y$ implies that $X \leq_{st} Y$ but, in general, the converse may not be true.

2.4. Convex ordering: Y is said to be larger than X in the convex ordering, denoted by $X \leq_{cx} Y$, if for every real-valued convex function $\phi(.)$defined on the real line, $E(\phi(X)) \leq E(\phi(Y))$.

2.5. Uniformly more variable ordering: X is said to be smaller than Y in uniformly more variable ordering, denoted by $X \leq_{uv} Y$, if X and Y have respective supports R_X and R_Y such that $R_X \subseteq R_Y$and the ratio $f(t)/g(t)$ is a unimodal function over R_Y but X and Y are not ordered in simple stochastic ordering. For random variables X and Y having a same mean, it is known that $X \leq_{uv} Y$ implies that $X \leq_{cx} Y$.

2.6. (Reversed) Hazard rate ordering: X is said to be smaller than Y in hazard rate ordering, denoted by $X \leq_{hr} Y$, if $r_Y(t) \leq r_X(t)$, for all $t \geq 0$. Similarly, X is

said to be smaller than Y in reversed hazard rate ordering, denoted by $X \leq_{rh} Y$, if $\bar{r}_X \leq \bar{r}_Y(t)$ for all $t \geq 0$. (Reversed) hazard rate ordering implies simple stochastic ordering but the converse may not be true. Also, $X \leq_{lr} Y$ is sufficient for $X \leq_{hr} Y$ and $X \leq_{rh} Y$.

2.7. Mean residual life ordering: X is said to be smaller than Y in mean residual life ordering, denoted by $X \leq_{mrl} Y$, if $\int_t^\infty \overline{F}(u)du / \int_t^\infty \overline{G}(u)du$ decreases in t, when defined. A sufficient condition for $X \leq_{mrl} Y$ is $X \leq_{hr} Y$.

2.8. (Factorial) Moments ordering: X is said to be smaller than Y in factorial moments ordering, denoted by $X \leq_{fm} Y$, if $E\binom{X}{i} \leq E\binom{Y}{i}$, for all $i \in \{1, 2, ...\}$. Similarly, X is said to be smaller than Y in moments ordering ($X \leq_{moments} Y$), if $E(X^i) \leq E(Y^i)$, for all $i \in \{1, 2, ...\}$. It is known that simple stochastic ordering implies factorial moments ordering and $X \leq_{fm} Y$ implies that $X \leq_{moments} Y$ and specially $X \leq_E Y$.

2.9. Moment generating function ordering: X is said to be smaller than Y in moment generating function ordering, denoted by $X \leq_{mgf} Y$, if $E(t^X) \geq E(t^Y)$ for all $t \in (0, 1)$. Simple stochastic ordering is sufficient for moment generating function ordering.

3. Comparison

Let X and Y be two r.v.'s with pmf's given in (1) and (2), respectively. To prove our main results, we first define:

$$a(k) = E(\frac{\Theta^{k+1}}{e^\Theta - 1}) / E(\frac{\Theta^k}{e^\Theta - 1}) \qquad k = 1, 2, \cdots \tag{3}$$

$$d_k(x) = x(\ln(1 + 1/x)^k)/k! \qquad , x > 0, k = 1, 2, ... \tag{4}$$

$$\lambda_0 = E\left(\frac{\Theta^2}{e^\Theta - 1}\right) / E\left(\frac{\Theta}{e^\Theta - 1}\right), \tag{5}$$

$$\lambda_1^* = \ln(1 + E(e^\Theta - 1)) \tag{6}$$

Then, we establish the following lemma.

Lemma 3.1 Consider notations (3) to (6). Then, we have

(a) $a(k)$ is a non-decreasing function in $k = 1, 2, \cdots$,

(b) for each fixed $k = 1, 2, \cdots$; $d_k(x)$ is a concave function in x>0,

(c) $s(x) = \frac{x}{e^x - 1}$ and $t(x) = \frac{xe^x}{e^x - 1}$are decreasing and increasing functions in $x > 0$, respectively.

Proof. (a) We may write for $a(k)$ in (3) that

$$a(k) = \int_0^\infty \frac{(\theta)^{k+1}}{e^\theta - 1} g(\theta) d\theta / \int_0^\infty \frac{(\theta)^k}{e^\theta - 1} g(\theta) d\theta = E(Z_k), \quad k = 1, 2, ...$$

where Z_k is a r.v. having pdf $h_k(z) = c_k \frac{z^k}{e^z - 1} g(z)$, for all $z \in (0, \infty)$ with c_k as the normalizing constant. Fixing $k \in \{1, 2, ...\}$, the ratio $h_{k+1}(z)/h_k(z)$is obviously a non-increasing function in $z \in (0, \infty)$, which implies that $Z_{k+1} \geq_{lr} Z_k$, and thus $Z_{k+1} \geq_{st} Z_k$. This, in turn, implies that $E(Z_{k+1}) \geq E(Z_k)$ or, equivalently, $a(k+1) \geq a(k)$. Since $k \in \{1, 2, ...\}$ was arbitrary, the assertion follows.

(b) It is sufficient to show that $d_j^{'}(x)$ is a decreasing function, for all $x > 0$. Since $\ln(1 + 1/x)$is a decreasing function, we can see that

$$d_j'(x) = \frac{1}{j!}(\ln(1+1/x))^j + jx(\frac{-1}{x^2})(\frac{1}{1+1/x})(\ln(1+1/x))^{j-1} \qquad j = 0, 1, 2, \cdots, k$$

is a decreasing function for all $x > 0$.

(c) It is trivially true, because $s'(x) < 0$ and $t'(x) > 0$, for all $x > 0$. □

The following theorem provides certain restrictions on the parameter λ to ensure that Y is larger than X in various orderings senses.

Theorem 3.1 Let X and Y be r.v.'s having distributions given by (1) and (2), respectively. Then, under the notations (3-6) and Lemma 3.1, we have

(a) $X \leq_{lr} Y$ if, and only if, $0 < \lambda \leq \lambda_0 = a(1)$,

(b) if $X \leq_{hr} Y$then $0 < \lambda \leq \lambda_1$ Conversely, if $0 < \lambda \leq \lambda_1^* \leq \lambda_1$ then, $X \leq_{hr} Y$.

(c) $X \leq_{rh} Y$ if, and only if, $0 < \lambda \leq \lambda_0 = a(1)$

(d) if $X \leq_{st} Y$ then $0 < \lambda \leq \lambda_1$. Conversely, if $0 < \lambda \leq \lambda_1^* \leq \lambda_1$ then, $X \leq_{st} Y$.

(e) $X \leq_E (\geq_E) Y$ if, and only if, $\lambda \leq (\geq) \lambda_2$.

Proof. (a) Consider the likelihood ratio

$$l(k) = \frac{\Pr(Y = k)}{\Pr(X = k)} = \frac{(e^\lambda - 1)}{\lambda^k} E\left[\frac{\Theta^k}{e^\Theta - 1}\right], \qquad k = 1, 2, ... \tag{7}$$

Then, by 2.3 we have

$$\begin{aligned} X \leq_{lr} Y &\Leftrightarrow l(k) \leq l(k+1) \qquad \forall\, k = 1, 2, \cdots \\ &\Leftrightarrow \lambda \leq \frac{E(\Theta^{k+1}/(e^\Theta - 1))}{E(\Theta^k/(e^\Theta - 1))} \quad \forall k = 1, 2, \cdots \\ &\Leftrightarrow \lambda \leq a(1) = \lambda_0 \quad \text{(by part (a) in Lemma 3.1)} \end{aligned}$$

which provides a necessary and sufficient condition for the ratio in (7) to be non-decreasing. Hence, the assertion (a) follows.

(b) First, let $X \leq_{hr} Y$. Then

$$\begin{aligned} &\frac{P(X = 1)}{P(x \geq 1)} \geq \frac{P(Y = 1)}{P(Y \geq 1)} \\ &\Rightarrow \frac{\lambda}{e^\lambda - 1} \geq E(\frac{\Theta}{e^\Theta - 1}) = \frac{\lambda_1}{e^{\lambda_1} - 1} \end{aligned}$$

Therefore, since $s(x) = \frac{x}{e^x-1}$ is a decreasing function, we have $\lambda \leq \lambda_1 = s^{-1}(E(s(\Theta)))$. Conversely, let $0 < \lambda \leq \lambda_1^*$ where λ_1^* is as in (6). For $k = 1, 2, \cdots$, consider

$$\begin{aligned}\Delta_\lambda(k) &= \frac{P(X=k)}{P(X\geq k)} - \frac{P(Y=k)}{P(Y\geq k)} \\ &= \frac{\lambda^k/(k!(e^\lambda-1))}{\sum_{j=k}^\infty \lambda^j/(j!(e^\lambda-1))} - \frac{E(\Theta^k/(k!(e^\Theta-1)))}{\sum_{j=k}^\infty E(\Theta^j/(j!(e^\Theta-1)))} \qquad k=1,2,\cdots\end{aligned}$$

It is easily seen that if $0 < \lambda \leq \lambda_1^*$, $\mathrm{X} \sim TP(\lambda)$ and $X_1 \sim TP(\lambda_1^*)$, then $X \leq_{lr} X_1$. Hence, by Ross (1983), we have $X \leq_{hr} X_1$. Consequently,

$$\begin{aligned}&\frac{P(X=k)}{P(X\geq k)} \geq \frac{P(X_1=k)}{P(X_1\geq k)} \\ \Rightarrow\ &\frac{\lambda^k/(k!(e^\lambda-1))}{\sum_{j=k}^\infty \lambda^j/(j!(e^\lambda-1))} \geq \frac{\lambda_1^{*k}/(k!(e^{\lambda_1^*}-1))}{\sum_{j=k}^\infty \lambda_1^{*j}/(j!(e^{\lambda_1^*}-1))}\end{aligned}$$

and thus for all k, $\Delta_\lambda(k) \geq \Delta_{\lambda_1^*}(k)$. Therefore, it is sufficient to show that $\Delta_{\lambda_1^*}(k) \geq 0$, $\forall k = 1, 2, \cdots$. Since

$$\begin{aligned}\frac{\lambda_1^{*k}}{k!(e^{\lambda_1^*}-1)} &= \frac{[\ln(1+E(e^\Theta-1))]^k}{k![e^{\ln(1+E(e^\Theta-1))}-1]} \\ &\geq \frac{1}{k!}E[\frac{(\ln(1+(e^\Theta-1)))^k}{(e^{\ln(1+(e^\Theta-1))}-1)}] \qquad \text{(by concavity of } d_k(.)) \\ &\geq \frac{1}{k!}E(\frac{\Theta^k}{e^\Theta-1})\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{j=k}^\infty \frac{(\lambda_1^*)^j}{j!(e^{\lambda_1^*}-1)} &= 1 - \sum_{j=1}^k \frac{(\ln(1+E(e^\Theta-1)))^j}{j!(e^{\ln(1+E(e^\Theta-1))}-1)} \\ &= 1 - \sum_{j=1}^k \frac{(\ln(1+E(e^\Theta-1)))^j}{j!(E(e^\Theta-1))} \\ &\leq 1 - \sum_{j=1}^k E(\frac{(\ln(1+(e^\Theta-1)))^j}{j!(e^\Theta-1)}) \qquad \text{(by concavity of } d_k(.)) \\ &= \sum_{j=k}^\infty E(\frac{\Theta^j}{j!(e^\Theta-1)})\end{aligned}$$

Therefore, $\Delta_{\lambda_1^*}(k) \geq 0, \forall k = 1, 2, \ldots$, as required. Note that $\lambda_1^* \leq \lambda_1$, because

$s(x) = \frac{x}{e^x-1}$ is a decreasing function and

$$\begin{aligned}\frac{\lambda_1^*}{e^{\lambda_1^*}-1} &= \frac{\ln(1+E(e^{\Theta}-1))}{e^{\ln(1+E(e^{\Theta}-1))}-1} \\ &\geq E(\frac{\Theta}{e^{\Theta}-1}) \quad \text{(by concavity of } d_1(.)) \\ &= \frac{\lambda_1}{e^{\lambda_1}-1}\end{aligned}$$

therefore, $\Delta_{\lambda_1}(k) \geq 0$, $\forall k = 1, 2, \cdots$, as required.
(c) First let $X \leq_{rh} Y$. Then $\frac{P(X=2)}{P(X\leq 2)} \leq \frac{P(Y=2)}{P(Y\leq 2)}$ which implies that $\lambda \leq \lambda_0$. Conversely, if $0 < \lambda \leq \lambda_0$ then, by part a, $X \leq_{lr} Y$ and consequently $X \leq_{rh} Y$.
(d) First, let $X \leq_{st} Y$. Then, $\Pr(Y \leq 1) \leq \Pr(X \leq 1)$ which implies that $E(\frac{\Theta}{e^{\Theta}-1}) = \frac{\lambda_1}{e^{\lambda_1}-1} \leq \frac{\lambda}{e^{\lambda}-1}$. Therefore, since $s(x) = \frac{x}{e^x-1}$ is a decreasing function, we have $\lambda \leq \lambda_1 = s^{-1}(E(s(\lambda)))$.
Conversely, suppose that $0 < \lambda \leq \lambda_1^*$ where λ_1^* is as in (6). Then, by part (b) above, $X \leq_{hr} X_1$ which results in $X \leq_{st} Y$.
(e) By definition 2.2, since $t(x) = \frac{x.e^x}{e^x-1}$ is an increasing function we have

$$\begin{aligned}X \leq_E Y &\Leftrightarrow E(X) \leq E(Y) \\ &\Leftrightarrow t(\lambda) = \frac{\lambda e^{\lambda}}{(e^{\lambda}-1)} \leq E[\frac{\Theta e^{\Theta}}{e^{\Theta}-1}] \\ &\Leftrightarrow \lambda \leq \lambda_2 = t^{-1}(E(t(\Theta))). \ \square\end{aligned}$$

The following corollary is valid by Theorem 3.1 and definitions 2.6, 2.7, 2.8 and 2.9.

Corollary 3.1
(a) for all $0 < \lambda \leq \lambda_0 = a(1)$, $X \leq_{mrl} Y$,
(b) for all $0 < \lambda \leq \lambda_1^*$, $X \leq_{fm} Y$, $X \leq_{mgf} Y$ and $X \leq_{moments} Y$,
(c) for all $\lambda > \lambda_1$, Y is not larger than X in simple stochastic and hence not in (reversed) hazard rate orderings.

Remark 3.1 Theorem 3.1 provides a condition under which sampling from a zero-truncated Poisson distribution is less favorable than that of its mixture.

In the next theorem, we shall establish that no value of $\lambda > 0$ can ensure that $Y \leq_{st} X$. Here we also make our comparison in the uniformly more variable ordering sense.

Theorem 3.2 Consider the notations of Theorem 3.1. Then, we have
(a) no value of $\lambda > 0$ can ensure that $Y \leq_{st} X$,
(b) if $X \leq_{uv} Y$, then $\lambda > \lambda_1^*$. Also, for $\lambda > \lambda_1$ we have $X \leq_{uv} Y$.

Proof. (a) It is proof by a similar argument used in Misra et al. (2003).
(b) Let $X \leq_{uv} Y$, then $\Pr(X = k)/\Pr(Y = k)$ is unimodal but X and Y are not ordered by the simple stochastic ordering. Thus, from Theorem 3.1(d) and part (a) above, it follows that $\lambda > \lambda_1^*$. Conversely, suppose that $\lambda > \lambda_1$, then from Theorem 3.1(d) and part (a) above, it is clear that r.v.'s Y and X are not ordered by the simple stochastic ordering. Also, from the arguments used in the proof of Theorem 3.1(a), it follows that $\Pr(X = k)/\Pr(Y = k)$ is unimodal, implying that $X \leq_{uv} Y$. □
Finally, as a special case, we compare the r.v.'s X and Y in (1) and (2) when they have a same mean. This is equivalent to the case $\lambda = \lambda_2$.

Theorem 3.3 Suppose that $\lambda = \lambda_2$. Then,
(a) $X \leq_{uv} Y$,
(b) $X \leq_{cx} Y$, which implies that $Var(X) \leq Var(Y)$,
(c) $\Pr(X \geq 2) > \Pr(Y \geq 2)$ and
(d) neither X is larger than Yin simple stochastic ordering sense, nor Y is larger than X.
Proof . (a) By Theorem 3.2 (b), since $\lambda_2 = \lambda > \lambda_1$, we have that $X \leq_{uv} Y$.
(b) Since $X \leq_{uv} Y$ and $E(X) = E(Y)$, we have $X \leq_{cx} Y$, i.e., $E(\phi(X)) \leq E(\phi(Y))$ for any convex function $\phi(.)$, such as $\phi(t) = t^2$. Thus $E(X^2) \leq E(Y^2)$, which implies that $Var(X) \leq Var(Y)$.
(c) Since $\lambda_2 = \lambda > \lambda_1$, and $s(x)$ is a decreasing function (Lemma 3.1(c)), we have

$$\frac{\lambda}{e^{\lambda} - 1} < \frac{\lambda_1}{e^{\lambda_1} - 1} = E(\frac{\Theta}{e^{\Theta} - 1}),$$

which yields $\Pr(X = 1) \leq \Pr(Y = 1)$, or equivalently $\Pr(X \geq 2) > \Pr(Y \geq 2)$.
(d) By part (a) above, we have$X \leq_{uv} Y$. So, Y is not larger than X in simple stochastic ordering sense. Also, by Theorem 3.2(a),X is larger than Y in simple stochastic ordering sense for no value of $\lambda > 0$. Therefore, when the zero-truncated Poisson random variable X has the same mean as that of its mixture Y, then there is no simple stochastic ordering between X and Y.□

Remark 3.2 Theorem 3.3 indicates that despite of lack of any simple stochastic ordering, when a zero-truncated Poisson distribution has the same mean as that of its mixture distribution, sampling from the zero-truncated Poisson distribution seems to be more favorable than its mixture.

Example In the sampling problem of different species, suppose that$X$, the number of moths per species caught in a light-trap, is a zero-truncated Poisson

r.v., TP(θ), with pmf

$$\Pr(X=k)=\frac{\theta^k}{k!(e^\theta-1)},\quad k=1,2,....$$

. It is interesting to specify whether it is better to take θ a random variable with some mixing distribution or to consider θ to be a fixed value in $(0,\infty)$. For simplicity of computation, assume that θ is an exponential random variable with unit mean. Then, obviously, Y, the number of individuals in different species, is a zero-truncated Poisson mixture r.v. with pmf

$$\Pr(Y=k)=\int_0^\infty \frac{\theta^k}{k!(e^\theta-1)}e^{-\theta}d\theta,\quad k=1,2,....$$

By Theorem 3.1, X is smaller than or equal to Y in likelihood ratio (reversed hazard rate) ordering sense if, and only if, $0<\lambda\leq\lambda_0=\frac{E\left(\Theta^2/(e^\Theta-1)\right)}{E(\Theta/(e^\Theta-1))}$. Since integration methods and numerical analysis for the computation of λ_0 are not applicable, we have used simulation runs of size $m=10000000$, and calculated $\lambda_0=0.626$ (it is rounded up to 3 digits). Also, by similar method, we calculated $\lambda_1^*=0.821, \lambda_1=0.693$ and $\lambda_2=1.094$ of (6) and Theorem 3.1. So, X is smaller than or equal to Y in simple stochastic (hazard rate) ordering sense, if $\lambda\leq\lambda_1=0.693$. In addition,$X$ is smaller than or equal to Y in expectation ordering sense if, and only if, $\lambda\leq\lambda_2=1.094$. Also, for $\lambda>0.626$ we have $X\leq_{uv}Y$(Theorem 3.2), specially, for $\lambda=1.094$, we conclude that $E(X)=E(Y)$, $X\leq_{uv}Y$, $X\leq_{cx}Y$, $Var(X)\leq Var(Y)$, $\Pr(X\geq 2)>\Pr(Y\geq 2)$ and there is no simple stochastic ordering between$X$ and Y (Theorem 3.3). □

References

[1] Aghababaei Jazi, M. and Alamatsaz, M.H. (2009): Ordering comparison of logarithmic series random Variables with their mixtures, *Communication in Statistics: Theory and Methods*, to appear.

[2] Alamatsaz, M.H. and Abbasi, S. (2008): Ordering comparison of negative binomial random Variables with their mixtures, *Statistics and Probability Letters*, Vol. 78, pp. 2234-2239.

[3] Campo, K., Gijsbrechts, E. and Nisol, P. (2003): The impact of retailer stockouts on whether, how much, and what to buy, *Marketing Research*, Vol. 20, pp. 273-286.

[4] Johnson, N.L. and Kotz, S. (1969): *Distributions in statistics- Discrete distributions*, John Wiley & Sons.

[5] Misra, N., Singh, H. and Harner, E.J. (2003): Stochastic comparisons of Poisson and binomial random variables with their mixtures, *Statistics and Probability Letters*, Vol. 65, pp. 279-290.

[6] Muller, A. and Stoyan, D. (2002): Comparison methods for stochastic models and risks, Wiley, StateplaceNew York.

[7] Ross, S.M. (1983): *Stochastic Processes*, Wiley, StateplaceNew York.
[8] Shaked, M. (1980): On Mixtures from Exponential Families, *J.Roy. Satist. Soc. Ser.*B 42, 192-198.
[9] Sapatinas, T. (1995): Identifiability of mixtures of power-series distributions and related characterizations, *Ann. Inst. Statist, Math*, Vol. 47, pp. 447-459.
[10] Shaked, M. and Shanthikumar, J.G. (2007): *Stochastic Orders*, Springer, StateplaceNew York.

Extensions of Stein's Lemma For the Unified Representation of Skewed Distributions

A. Abtahi[1] and M. Towhidi[2]

[1] *Department of Mathematics, Shiraz Islamic Azad University*
[2] *Department of Statistic, Shiraz University*

In this paper we consider a constructive representation of skewed distributions, which proposed by Ferreira and Steel (2006), and its basic properties is presented. We show that every continuous skewed distribution can be stated as this unified form. In order to illustrate the effect of different skewing mechanism, we consider four mechanisms with different degrees of skewness and compare with each other. The main aim of this paper is to present stein's lemma for the new unified skew-normal distribution. This work extends Adcock (2007), which present two versions of stein's lemma for skew-normal distribution.

Keywords: Skewed Distribution; Inverse Probability Integral Transformation, Skew-Normal; Beta-normal Density; Inverse Scale factor Density; Stein's Lemma.

1. Introduction

The first systematic treatment of the skew-normal (SN) class of distributions was given by Azzalini (1985). Subsequently, Azzalini and Dalla Valle (1996) introduced a multivariate version of the skew-normal density. Developments and generalizations of these ideas to various directions and for other distributions have been proposed by many authors. A review of these research works can be found the paper of Azzalini (2005)and the book edited by Genton (2004). There are so many proposals in construction skewed distributions and it is worth to find an overall class which covers all of these proposals. Recently, Ferreira and Steel (2006) proposed a general perspective on the skewing a symmetric distribution. The idea was to separate the skewing mechanism from a symmetric distribution and the representation of skewed distributions was based on the inverse probability integral transformations. This unified form covers a wide set of classes of skewed distributions. In this paper, we show that every continuous skewed distribution can be stated with this unified form.

The main aim of this paper is to present stein's lemma for the unified skew-normal distribution. This work extends Adcock (2007), which present two ver-

sions of stein's lemma for skew-normal distribution. As summarized in Adcock (2007), Stein's lemma plays an important role in modern portfolio theory. The theory of portfolio selection developed by Harry Markowitz, see for example Markowitz(1952,1987), asuumes that investors mimimize the variance of portfolio returns subject to achieeving a given target expected return. It is well known that returns on financial assets are not normally distributed. They exhibit both skewness and kurtosis. Therfore, we extend the stein's lemma for the unified skew-normal distribution.

The structure of this paper is as follows. In Section 2 we define the representation of unified skewed distributions and its basic properties. The section briefly reports four classes of skewed distributions that are required in the rest of the paper. Section 3, presents the main results of the paper, stein's lemma for the unified of skew-normal distribution. This section also contains corollaries to the result, including siegel's formula.

2. Unified skewed distribution, $USD(f, p)$

In this section, we introduce the representation of unified skewed distribution and prove a number of properties based on this representation.

A random variable $X_{f,p}$ is said to have a unified skewed distribution with functional parameters f and p, if its *pdf* is of the form

$$s(y|f,p) = f(y)\, p[F(y)]; \qquad -\infty < y < \infty, \tag{1}$$

where f is a symmetric density function on the real line and p is a *pdf* on (0,1). In fact the density p specifies a skewing mechanism. We denote a random variable with this unified skewed distribution by $X_{f,p} \sim USD\,(f,p)$.

In order to illustrate the effect of different skewing mechanisms, we compare four mechanisms with different degrees of skewness.

2.1. *Comparison of the four versions of unified skewed distributions*

For comparing the effect of different skewing mechanisms on a symmetric distribution, we consider the following four mechanisms of skewing normal distribution.

(i) Hidden truncation method : The most common version of univariate skewed distributions, can be generated by some hidden truncation developments which involve the densities of the form:

$$s(y) = 2f(y)G(w(y)), \tag{2}$$

where G denotes the *cdf* of a symmetric distribution on $\Re$, f is a symmetric *pdf* on $\Re$, and $w(y)$ is an odd function. (For more details, see Arnold and Beaver (2002) and Azzalini and Capitanio (2003)).

In the model (2), by defining the *pdf* p as, $p\left(y\right)=2G(w(F^{-1}(y)))$, we can interpret the model (2) in the form (1). Two well known family of skewed symmetric distributions, in this class, were introduced by Azzalini (1985) and Arellano-Valle, et.al (2004) as follows:

$$s_1\left(y\left|\lambda\right.\right)=2\phi\left(y\right)\Phi\left(\lambda\, y\right),\ \ and\ \ \ s_1^*(y|\lambda_1,\lambda_2)=2\phi(y)\Phi(\frac{\lambda_1 y}{\sqrt{1+\lambda_2 y^2}});\quad y\in\Re, \tag{3}$$

where $\phi\left(y\right)$ and $\Phi\left(y\right)$ are density and distribution functions of a standard normal random variable, respectively. Figure 1 shows the shape of $s_1^*(y|\lambda_1,\lambda_2)$ and the skewing mechanism $p_1^*(y|\lambda_1,\lambda_2)=2\Phi[\frac{\lambda_1\Phi^{-1}(y)}{\sqrt{1+\lambda_2(\Phi^{-1}(y))^2}}]$ for some values of $\lambda_1$and λ_2 . Azzalini's density (or mechanism) occurs if $\lambda_2=0$.

(ii) Order statistics method: This method contains a family of skewed distributions, which is generated by taking F as the standard normal distribution and P as the Beta distribution with unknown parameters a and b, in the $USD\left(f,p\right)$.

These distributions have some properties that can make them less attractive in applications. Eugen et al. (2002) mentioned that the Beta-Normal distribution can be bimodal and the amount of skewness is not a monotone function of a and b and is not flexible enough to highly skewed distributions. Ferreira and steel (2004b) suggested a restricted parameterization on the Beta-Normal distribution by taking $b=\frac{1}{a}$, which leads to a unimodal and monotone density function. Therefore, we consider the following form of the *pdf* with a real parameter $a>0$, as the second skewed distribution.

$$s_2\left(y\left|a\right.\right)=\frac{1}{B\left(a,\frac{1}{a}\right)}\,\phi\left(y\right)\,\left[\Phi\left(y\right)\right]^{a-1}\,\left[1-\Phi\left(y\right)\right]^{\frac{1}{a}-1};\qquad y\in\Re. \tag{4}$$

The above distribution is always skewed for $a\neq 1$ and the values of a larger (smaller) than one, correspond to positively (negatively) skewed distribution. This model can be interpreted as model (1) by defining p_2 as a Beta distribution. The connection between parameter a and the skewness of *pdf* s_2 and p_2 is shown in Figure 1.

(iii)Inverse scale factors method: A family of skew-normal distributions generated by introducing scale factors in positive and the negative half real lines. A general form of this distribution is introduced by Fernandez and Steel (1998). This distribution has density

$$s_3(y|\gamma)=\frac{2}{\gamma+\frac{1}{\gamma}}\phi[y\gamma^{-sign(y)}];\qquad y\in\Re, \tag{5}$$

with $sign(.)$ as the usual $sign$ function in $\Re$, and γ is a scaler in $(0,\infty)$. If $\gamma>1$, then the distribution is right-skewed, whereas it is left-skewed for $\gamma<1$. This

distribution can be constructed by model (1) by choosing $f = \phi(.)$ and using p_3 as $p_3(y|\gamma) = \frac{2}{\gamma+\frac{1}{\gamma}}.\frac{\phi[\gamma^{sign(\frac{1}{2}-y)}\Phi^{-1}(y)]}{\phi[\Phi^{-1}(y)]}$. Figure (1) illustrates several possible shapes obtained from $s_3(.)$ and $p_3(.)$ under various choices of γ.

(iv)Unified method: A larger class of skewed distributions can be constructed by defining F as the standard normal distribution and taking an arbitrary *pdf* p on $(0, 1)$. In this case, the *pdf* is given by

$$s_4(x|p) = \phi(x)\, p[\Phi(x)]; \qquad y \in \Re. \tag{6}$$

An illustration of the effect of this skewness mechanism is shown in Figure 1, by assuming a triangular density function for $p\,(.)$ as:

$$p_4(y|\alpha) = \begin{cases} \frac{2}{\alpha}y & \text{if } 0 < y < \alpha \\ \frac{2}{(1-\alpha)}(1-y) & \text{if } \alpha < y < 1 \end{cases}$$

where $\alpha \in (0, 1)$.

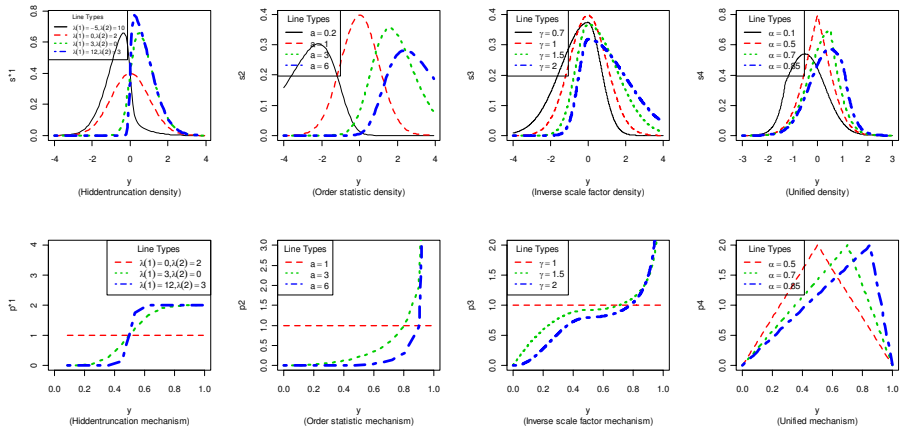

Fig. 1. Comparison of the four versions of unified skewed distributions

All of these methods skew a symmetric distribution F but in different ways. Indeed, any skewed version of the same symmetric distribution can then be modeled directly by choosing a particular p. For the hidden truncation technique, there exists a considerable amount of skewness around $(\frac{1}{2})$. But the beta-mechanism shows more various skewness in the tail of the distribution $p(.)$. In the inverse scale factors method, we see that by varying y from zero to unity, $p(y)$ changes moderately and large values of skewness give fat tail to $s(.)$. The especial case, in

the illustration of the method (iv) shows that a positive (negative) skewness on $p(.)$ yields a positive (negative) skewness on $s(.)$.

2.2. *Some representation and properties of* $USD(f,p)$

In Section 2.2, we discussed how various versions of skew-normal distribution arise with different skewing mechanism. The following theorem states that every continuous skewed distribution can be stated in the form of unified skewed distribution in (1).

Theorem 1 *Every continuous skewed density function* $s(.)$ *can be stated in the form of* $USD(f,p)$.

Proof. Let s and S be the *pdf* and *cdf* of a skewed distribution, respectively. Now we have to find a symmetric *pdf* f on the real line and a *pdf* p on (0,1) such that $s(y)$ can be written as:

$$s(y) = f(y)p\ [F(y)]. \tag{7}$$

If, we define a symmetric *pdf* f as:

$$f(y) = \frac{s(y) + s(-y)}{2}, \tag{8}$$

then, the *cdf* corresponding to f is in the form:

$$F(y) = \tfrac{S(y)+(1-S(-y))}{2}$$

and from (7) the *pdf* p on (0,1) is:

$$p\left(F\left(y\right)\right) = \frac{2\,s\left(y\right)}{s\left(y\right) + s\left(-y\right)}$$

or

$$p(y) = \frac{2\,s(F^{-1}(y))}{s(F^{-1}(y)) + s(-F^{-1}(y))}$$

It is enough to show that $p\left(y\right)$ is a *pdf* on (0,1). Note that

$$\int_0^1 p\left(y\right)\,dy = \int_{-\infty}^{+\infty} \frac{2\,s\left(t\right)}{s\left(t\right) + s\left(-t\right)}\,f\left(t\right)\,dt = 1$$ □

Some applications of the above theorem are given below.

(a) Some selection mechanisms lead to many existing families of skewed distributions, such as skew-normal and skew-elliptical distributions. This class is defined by Arrellano-Valle et al (2006) with the following function:

$$s(y) = f(y)\frac{P[Z \in S \mid Y = y]}{P[Z \in S]} = f(y)\frac{Q(y)}{E[Q(Y)]} \tag{9}$$

where f denotes the marginal *pdf* of Y and $Q(y) = P[Z \in S \mid Y = y]$, with random vector (Y, Z) and selection set S. If f is a symmetric density and define *pdf* p as,

$$p(F(y)) = \frac{P[Z \in S | U = F(y)]}{P[Z \in S]},$$

where $U = F(Y) \sim U(0, 1)$, or

$$p(y) = \frac{P[Z \in S | Y = F^{-1}(y)]}{P[Z \in S]},$$

then the density (9) is equivalent to (1).

(b) The class of skewed-elliptical distributions which was introduced by Genton and Loperfido (2005), in the form $s(y) = 2f(y)\pi(y)$, with a symmetric *pdf* f and the skewing function π, can be interpreted in the form (1) by choosing the *pdf* p as $p(y) = 2\pi\ [F^{-1}(y)]$.

(c) The generalized skew-normal distributions in Gupta and Gupta (2004) with density $s(y) = \frac{\phi(y)\Phi^n(\lambda y)}{c_n(\lambda)}$, can be introduced by choosing $f = \phi$, and $p(y) = \frac{\Phi^n[\lambda\Phi^{-1}(y)]}{c_n(\lambda)}$.

Some interesting representations and properties of the $USD\,(f, p)$ are given in the following propositions.

Proposition 2.1. *Let U and V be two independent random variables with pdf s (cdf) f (F) on the real line and p on (0, 1), respectively. Then the following results are concluded.*
(a) *$F^{-1}\,(V) \sim USD\,(f, p)$*
(b) *If $W = V - F(U)$, then the conditional distribution of U given $(W = 0)$ is $USD\,(f, p)$.*
(c) *If $p(1-y) = 2-p(y)$, then $T = U.S_U \sim USD(f, p)$, where conditionally on $U = u$, $S_U = 1$ with probability $\frac{1}{2}p[F(u)]$ and $S_U = -1$ with probability $\frac{1}{2}[2 - p[F(u)]]$.*

Proof. Part (a) is due to Ferreira and Steel (2006) and the proof of part (c) is obvious. For a proof of part (b), note that the conditional density function of U given $(W = 0)$ is

$$\begin{aligned} f_{U|W}(u|0) &= \frac{f_{U,W}(u, 0)}{f_W(0)} \\ &= \frac{f_U(u) f_V(F(u))}{\int_{-\infty}^{\infty} f_U(u) f_V(F(u)) du} \\ &= \frac{f(u)p(F(u))}{\int_{-\infty}^{\infty} f(u)p(F(u)) du} \\ &= f(u)p(F(u)). \end{aligned}$$

□

Proposition 2.2. *Let U and V be two independent random variables with pdf s f, p and cdf s F, P, respectively. Also, let $X_{f,p} \sim USD(f,p)$. Then the following results are concluded*

(a) *If p is symmetric around $\frac{1}{2}$, then $X_{f,p}$ will be a symmetric random variable.*

(b) *When P is the uniform distribution on (0,1), then $X_{f,p} \overset{d}{=} U$.*

(c) *$F(X_{f,p}) \overset{d}{=} V$.*

(d) *$-X_{f,p} \overset{d}{=} X_{f,p^*}$, where $p^*(y) = p(1-y)$; $0 < y < 1$.*

(e) *$|X_{f,p}| \overset{d}{=} |U|$, if $p(1-y) = 2 - p(y)$.*

(g) *$X^2_{\phi,p} \overset{d}{=} \chi^2(1)$, if $p(1-y) = 2 - p(y)$ where ϕ denotes the standard normal density.*

(h) *The skewness of $X_{f,p}$ is positive (negative), if the skewness of the pdf p is positive(negative).*

(i) *The r-th moment of $X_{f,p}$ with $USD(f,p)$ depends on the density functions f, p and can be evaluated by the relation:*

$$E[X^r] = E\{[F^{-1}(V)]^r\}.$$

3. Extension of stein's lemma for the unified of skew-normal distribution

When a random variable X has a normal distribution, stein's lemma (stein, 1981), states that, under certain regularity conditions on a function g(.)

$$Cov[X, g(X)] = Var(X)E[g'(X)]. \tag{10}$$

A generalization of this result is given by Liu(1994), Adcock(2007) and Landsman and Neslehova (2008).

We extend the above result and prove stein's lemma for the unified skew-normal distribution.

Theorem 3.1. *Let X be a random variable distributed as $USD(\phi,p)$; where ϕ is pdf of a normal distribution with mean μ and variance σ^2. Let $h(x)$ be a continuous real valued function such that $h'(x)$ exist and $E[h'(X)] < \infty$. Then, we have*

$$Cov[X, h(X)] = \sigma^2\{E[h'(X)] + E[h(Z)\frac{d}{dZ}p^*[Z] - E[\frac{d}{dZ}p^*[Z]]E[h(X)]\} \tag{11}$$

where $Z \sim N(\mu, \sigma^2)$ and $p^*[Z] = p[\Phi(Z)]$.

Proof.

$$Cov[X, h(X)] = E[Xh(X)] - E[X]E[h(X)] \tag{12}$$

and

$$\begin{aligned} E[Xh(X)] &= \int_{-\infty}^{+\infty} xh(x)\phi(x)p[\Phi(x)]dx \\ &= E[Zh(Z)p^*(Z)] \\ &= E[Zg(Z)] \end{aligned} \tag{13}$$

where $g(z) = h(z)p^*(z)$, $p^*(z) = p[\Phi(z)]$ and $Z \sim N(\mu, \sigma^2)$.

By Stein's lemma in (10), we have $E[Zg(Z)] = \sigma^2 E[g'(Z)] + \mu E[g(Z)]$. Now note that

$$E[g(Z)] = E[h(Z)p^*(Z)] = E[h(X)] \tag{14}$$

and

$$\begin{aligned} E[g'(Z)] &= E[h'(Z)p^*(Z)] + E[\frac{d}{dZ}p^*(Z)h(Z)] \\ &= E[h'(X)] + E[\frac{d}{dZ}p^*(Z)h(Z)]. \end{aligned} \tag{15}$$

Combining the relations (13)-(15) lead us to the following equation

$$E[Xh(X)] = \sigma^2 E[h'(X)] + \sigma^2 E[\frac{d}{dZ}p^*(Z)h(Z)] + \mu E[h(X)]. \tag{16}$$

On the other hand $E[p^*(Z)] = 1$, so

$$\begin{aligned} E[X] = E[Zp^*(Z)] &= \sigma^2 E[\frac{d}{dZ}p^*(Z)] + E[Z]E[p^*(Z)] \\ &= \sigma^2 E[\frac{d}{dZ}p^*(Z)] + \mu. \end{aligned} \tag{17}$$

Now using (12), (16) and (17), the result follows and the proof is complete. □

Theorem 3.1 leads in turn to a generalization of Siegel's formula (1993) for the covariance of an arbitrary variable of unified skew-normal distribution with its i-th order statistics.

Corollary 3.1. *Let $X_1, ..., X_n$ be a random sample from $USD(\phi, p)$, then*

$$Cov[X_j, X_{(i)}] = \sigma^2 P[X_j = X_{(i)}] + E[\frac{dp^*[Z]}{dZ}.Z_{(i)}] - E[\frac{d}{dZ}p^*[Z]]E[Z_{(i)}] \tag{18}$$

Proof. Define $h(x) = x_{(i)}$, where $x_{(i)}$ is the i-th largest variable among the random sample of $USD(\phi, p)$. The derivatives of $h(.)$ with respect to the $j - th$ random variable

$$\frac{\partial h}{\partial x_j} = \begin{cases} 1 \text{ if } x_j = x_{(i)} \\ 0 \text{ otherwise} \end{cases}$$

then Theorem 3.1 yields the result. □

In the following corollary, we derive Stein's identity in especial cases.

Corollary 3.2. *Let X be a random variable with $USD(\phi, p)$, ϕ is the pdf of a normal distribution with mean 0 and variance σ^2.*
(a) If $p(1-x) = 2 - p(x)$, and $h(.)$ is a real-valued odd function for all $x \in \Re$, then

$$Cov[X, h(X)] = \sigma^2\{E[h'(X)] - E[\frac{d}{dZ}p^*[Z]]E[h(X)]\} \tag{19}$$

(b)If X has a skew-normal distribution (3),$s_1(.|\lambda)$, then for any real valued function $h(.)$

$$Cov[X, h(X)] = \sigma^2 E[h'(X)] + \mu_{SN}(E[h(Z^*)] - E(h(X))) \tag{20}$$

where $Z^* \sim N(0, (1+\lambda^2)^{-\frac{1}{2}})$ and μ_{SN} is the mean of the random variable X.

Proof. (a) By using the fact that $h(-x) = -h(x)$ and $\frac{d}{dx}p(1-x) = \frac{d}{dx}p(x)$ we conclude that $E[h(Z)\frac{d}{dZ}p[\Phi(Z)]] = 0$. Therefore, the result follows.

(b) In this case $\frac{d}{dz}p^*(z) = \frac{d}{dz}2\Phi(\lambda z) = \lambda\phi(\lambda z)$, therefore $E[h(Z)\frac{d}{dZ}p^*(Z)]$ and $E[\frac{d}{dZ}p^*(Z)]$ are obtained as:

$$\begin{aligned} E[h(Z)\frac{d}{dZ}p^*(Z)] &= 2\lambda \int h(z)\phi(z)\phi(\lambda z)dz \\ &= (\frac{2}{\pi})^{\frac{1}{2}}\frac{\lambda}{\sqrt{1+\lambda^2}}\int \frac{\sqrt{\lambda^2+1}}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2(\lambda^2+1)}dz \\ &= \mu_{SN}E[h(Z^*)] \end{aligned}$$

where $Z^* \sim N(0, (1+\lambda^2)^{-\frac{1}{2}})$ and, by Azzalini (1985), $\mu_{SN} = (\frac{2}{\pi})^{\frac{1}{2}}\frac{\lambda}{\sqrt{1+\lambda^2}}$ is the mean of the skew-normal distribution.

Similarly, we have $E[\frac{d}{dZ}p^*(Z)] = \mu_{SN}$ and the result is obtained by Theorem 3.1. □

At the end, it is necessary to note that some properties of USD, stated in this paper, are fulfilled just under the condition $p[1-x] = 2 - p[x]$. In many families of skewed distributions such as hidden truncation density, and skew-elliptical which have been introduced in (2) and (2.2 (b)), the skewing mechanism p has such a condition, which can be verify easily.

References

1. C. J. Adcock, Extensions of Stein's lemma for the skew- normal distribution, *Communications in statistics-Theory and Methods* **36**, 1661-1671 (2007).

2. R. B. Arellano-Valle, M. D. Branco and M.G. Genton, A unified view on skewed distributions arising from selections, *Canadian Journal of Statistic* **34**, 581-601 (2006).
3. R. B. Arellano-Valle, H. W. Gomez and F. A. Quintana, A new class of skew-normal distributions, *Communications in statistics-Theory and Methods* **33(7)**, 1465-1480 (2004).
4. B. C. Arnold and R. J. Beaver, Skewed multivariate models related to hidden truncation and / or selective reporting (with discussion.) *Test II* 7-54 (2002).
5. A. Azzalini, A class of distributions with includes the normal ones, *Scandinavian Journal of Statistics* **12**, 171-178 (1985).
6. A. Azzalini, The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics* **32(2)**, 159-188 (2005).
7. A. Azzalini and A. Capitanio, Distributions generated by perturbation of symmetry-with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society Series B* **65**, 367-389 (2003).
8. A. Azzalini and A. Dalla Valle, The multivariate skew-normal distribution, *Biometricka* **83(a)**, 715-726 (1996).
9. N. Eugene, C. Lee and F. Famoye, Beta-normal distribution and its applications, *Communicational in Statistics: Theory Methods* **31**, 497-512 (2002).
10. J. T. A. S. Ferreira and M.F.J. Steel, Discussion of MC Jones, Families of distributions arising from distributions of order statistics, *Test* **13**, 30-33 (2004b).
11. J. T. A. S. Ferreira and M.F.J. Steel, A constructive representation of univariate skewed distributions, *Journal of the American Statistical Association* **101**, 823-829 (2006).
12. C. Fernandez and M. F. J. Steel, On Bayesian modeling of fat tails and skewness, *Journal of the American Statistical Association* **93**, 359-371 (1998).
13. M. G. Genton (ed.) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality CRC* (Chapman and Hall, London, Boca Raton, FL (2004)).
14. M. G. Genton , N. Loperfido, Generalized Skew-Elliptical Distributions and Their Quadratic Forms, *Annals of the Institute of Statistical Mathematics* **57**, 389-401 (2005).
15. R.C. Gupta and R.D. Gupta, Generalized Skew Normal Model, *Test* **13(2)**, 501-524 (2004).
16. J. S. Liu, Siegel's formula via Stein's identities, *Statistics and probability letters* **21**, 247-251 (1994).
17. Z. Lanssman and J. Neslehova, Stein's lemma for elliptical class of distributions, *Statistics and Probability Letters* **76**, 912-927 (2008).
18. H. M. Markowitz, Portfolio selection, *J. Fin.* **7**, 77-91 (1952).
19. H. M. Markowitz, Mean variance Analysis in Portfolio Choice and Capital Markets, (oxford: Blackwell (1987)).
20. A. F. Siegel, A surprising covariance involving the minimum of multivariate normal Variables, *Journal of American statistical association* **88**, 77-80 (1993).
21. C. M. Stein, Estimation of the mean of a multivariate normal distribution, *Ann. Statist.* **9**, 119-35-1151 (1981).

Some Extensions of Discrete α-Monotone Distributions

M. Aghababaei Jazi and M.H. Alamatsaz

Faculty of Mathematics, University of Sistan and Balochestan

Faculty of Sciences, University of Isfahan

Binomial thinning operator was introduced to define α-monotone distributions on the lattice of integers. This operator is actually a compound of i.i.d. Bernoulli random variables which was also employed to define self-decomposable distributions on the lattice of nonnegative integers. Several other different thinning operators have been defined to model discrete data, specially, in integer-valued time series. In this paper, we shall introduce two new thinning operators based on zero-inflated and inflated-parameter Bernoulli random variables. Then, based on these, we shall define modified and generalized α-monotonicity and investigate their similar properties and characterizations.

Keywords: Binomial thinning operator; Compound distributions; Inflated-parameter distributions; Zero-inflated distributions.

1. Introduction

Thinning operators are probabilistic operations that can be applied to random counts. The basic idea is that count represents the random size of an imaginary population, and the thinning operation randomly deletes some members of this population. Since the size of the shrunken population is still integer-valued, the population of thinning always leads to integer values.

The first and most popular thinning operation is binomial thinning operator, defined by Steutel and van Harn [1] as follows:

Definition 1.1. Let Y be an arbitrary nonnegative integer-valued random variable (r.v.) and $X_1, X_2, ...$ be i.i.d. Ber(u), $u \in [0, 1)$. Then

$$u \otimes Y = \sum_{i=1}^{Y} X_i. \tag{1}$$

is called the binomial thinning operator(or u-fraction) of Y.

For the interpretation of the binomial thinning operation, consider a population of size X at a certain time t. If we observe the same population at a later point

of time, say $t+1$, then the population may be shrunken, because some of the individuals died between times t and $t+1$. If the individuals die independently of each other, and if the probability of dying in between t and $t+1$ is equal to $1-u$ for all individuals, then the number of survivors is given by $u \otimes X$.
Since binomial thinning always leads to an integer and since both $u \otimes X$ and $u.X$ have the same mean, the scalar multiplication of ARMA recursion was replaced by binomial thinning in the definition of integer-valued ARMA (INARMA) models. See, e.g., [2] on INAR(1) model as $X_t = \alpha \otimes X_{t-1} + \epsilon_t$.

Based on operator $\otimes$, Steutel [3] defined a discrete analogue of α-monotonicity of Olshen and Savage [4] as follows.

Definition 1.2. A nonnegative integer-valued r.v. X is α-monotone, if

$$X =^d U^{1/\alpha} \otimes Y. \tag{2}$$

where Y is a nonnegative integer-valued r.v. and $U \sim U(0,1)$, independent of Y, where $=^d$ means equality in distribution.

He showed that

Theorem 1.1. *(Steutel [3]) A nonnegative integer-valued r.v. X with distribution $\{p_n\}_0^\infty$ and probability generating function (pgf) P_X, is α-monotone if, and only if (iff)*

$$P_X(t) = \alpha(1-t)^{-\alpha} \int_t^1 (1-w)^{\alpha-1} Q(w) dw,$$

where $Q(.)$ is the pgf of a nonnegative discrete r.v.. Or, equivalently, iff

$$(\alpha+n)p_n \geq (1+n)p_{n+1}, \qquad \forall n \geq 0\,.$$

According to Steutel [3], a discrete distribution $\{p_n\}_{-\infty}^\infty$ is called α-unimodal (about zero) if for some $\alpha > 0$,

$$\begin{cases} (\alpha-n)p_n \geq (1-n)p_{n-1}\,, n \leq 0 \\ (\alpha+n)p_n \geq (1+n)p_{n+1}\,, n \geq 0\,. \end{cases}$$

Obviously, 1-unimodality is just the usual unimodality.

Alzaid and Al-Osh [5] gave a similar characterization to that of Olshen and Savage [4] for discrete α-monotone distributions as follows.

Theorem 1.2. *(Alzaid and Al-Osh [5]) A nonnegative integer-valued r.v. X is α-monotone iff for every nonnegative bounded measurable function g, $t^\alpha E[g(t \otimes X)]$ is non-decreasing in $t \in (0,1]$.*

Also, Steutel and van Harn [1] used binomial thinning operator to define a discrete self-decomposable r.v.. They called a nonnegative discrete r.v. X self-decomposable if for any $\alpha \in (0,1)$ there exists a r.v. X_α independent of X such that $X =^d \alpha \otimes X + X_\alpha$ (see [6] for more details).

Binomial thinning operator is quite attractive, not only because of its mathematical elegance, but also because of its quite intuitive interpretation. Nevertheless, it becomes clear in the practice of time series modeling that binomial thinning is best suited for Poisson marginals. Therefore, several alternatives to binomial thinning have been defined which can be applied to distributions other than Poisson (see, e.g., Weiβ [7]). In this paper, we shall introduce two new thinning operators based on zero-inflated and inflated-parameter Bernoulli r.v.'s. Then, based on these, we shall define modified and generalized α-monotonicity and study their similar properties and characterizations.

Let ξ be an arbitrary nonnegative integer-valued r.v. such that $Pr(\xi = j) = p_j$, $j = 0, 1, 2, ...$ and $\Sigma_{j=0}^{\infty} p_j = 1$. If an extra proportion of zeros, $\rho \in [0,1]$, is added to zeros of the r.v. ξ, while decreasing the remaining proportions in an appropriate way, the zero-inflated modification η of ξ is defined by

$$Pr(\eta = j) = \begin{cases} \rho + (1-\rho)p_0 & , j = 0 \\ (1-\rho)p_j & , j = 1, 2, \dots . \end{cases}$$

It is clear that if $\rho = 1$, the corresponding zero-inflated distribution is degenerate at zero and if $\rho = 0$, there is no inflation, i.e., $\eta =^d \xi$.

For most cases, the inflating parameter ρ lies between 0 and 1, although it may also take negative values, provided that $P(\eta = 0) > 0$, i.e., $\rho \geq -\frac{p_0}{1-p_0}$. This latter case corresponds to the opposite phenomena, i.e., excludes a proportion of zeros from the basic discrete distribution. In this paper, we consider the former case, where $\rho \in [0,1)$. For example, the distribution of a zero-inflated Bernoulli r.v. X with parameter $u \in (0,1)$ and the inflating parameter ρ, denoted by $X \sim$ZIBer(u, ρ), is given by

$$Pr(X = j) = \begin{cases} \rho + (1-\rho)(1-u) & , j = 0 \\ (1-\rho)u & , j = 1 . \end{cases} \tag{3}$$

Based on a sequence of ZIBer(u, ρ) r.v.'s, Kolev et al [8] introduced inflated-parameter modification of geometric distribution. They showed that the pgf of an inflated-parameter geometric r.v. S, denoted by IGe(π, ρ), is given by

$$P_S(t) = \frac{\pi(1 - t\rho)}{1 - t[1 - \pi + \rho\pi]},$$

which is actually the pgf of $S = X_1 + X_2 + ... + X_N$, where N is a geometric r.v. with parameter $\pi \in (0,1)$ and pgf $P_N(t) = \frac{\pi}{1-(1-\pi)t}$ in $t \in [0,1]$, denoted by Ge$_0(\pi)$. Also, $X_1, X_2, ..., X_N$ are geometric r.v.'s with parameter $1 - \rho$ and

pgf $P_X(t) = \frac{t(1-\rho)}{1-t\rho}$ in $t \in [0,1]$, denoted by $\mathrm{Ge}_1(1-\rho)$, independent of r.v. N. Similarly, if N is a r.v. distributed according to the binomial, negative binomial or Poisson distributions, then $S = X_1 + X_2 + ... + X_N$ is inflated-parameter binomial, negative binomial or Poisson r.v., respectively. For example, the pmf of a inflated-parameter Bernoulli r.v. X with parameter $u \in (0,1)$ and the inflating parameter $\rho \in [0,1)$, denoted by $\mathrm{IBer}(u,\rho)$, is given by

$$Pr(X=j) = \begin{cases} 1-u & , j=0 \\ u\rho^{k-1}(1-\rho) & , j=1,2,.... \end{cases} \tag{4}$$

For more details on zero-inflated and inflated-parameter Bernoulli distributions, we refer the readers to Kolev et al [8] and Minkova [9].

In section 2, we shall modify the binomial thinning operator $\otimes$ using zero-inflated Bernoulli r.v.'s and discuss their properties. Section 3 concerns with a generalization of the concept using inflated-parameter Bernoulli r.v.'s. In both cases, we shall discuss their convolution properties and provide some characterization results.

2. ρ-Modified α-Monotone Discrete Distributions

We shall first consider an extension of the binomial thinning operator $\otimes$ introduced by Steutel and van Harn [1], based on zero-inflated Bernoulli r.v.'s. Then, accordingly, we shall introduce a modification of α-monotonicity.

Definition 2.1. Let Y be a discrete r.v. on N_0 and $X_1, X_2, ...$ be i.i.d. $\mathrm{ZIBer}(u,\rho)$ r.v.'s with $\rho, u \in [0,1)$. Then, we call the operator $\otimes^\rho$ defined by

$$u \otimes^\rho Y = \sum_{i=1}^{Y} X_i \tag{5}$$

a zero-inflated thinning operator with inflating parameter ρ.

In the case $\rho = 0$ (no inflation), zero-inflated thinning operator in (5) is just the (usual) binomial thinning operator. Further, we can see that

$$u \otimes^\rho Y =^d u(1-\rho) \otimes Y \tag{6}$$

because, let P_Y be the pgf of Y, then pgf of $u \otimes^\rho Y$ is

$$\begin{aligned} P_{u\otimes^\rho Y}(t) &= EE(t^{u\otimes^\rho Y}|Y) \\ &= EE(t^{\sum_{i=1}^{Y} X_i}|Y) \\ &= E[E(t^{X_1})]^Y \\ &= E[\rho + (1-\rho)(1-u+ut)]^Y \qquad (\text{from } (3)) \\ &= P_Y(1-u(1-\rho)+u(1-\rho)t) \\ &= P_Y(1-\upsilon+\upsilon t) \\ &= P_{\upsilon\otimes Y}(t)\,, \end{aligned}$$

where $\upsilon = u(1-\rho)$. Therefore, the two operators $\otimes$ and $\otimes^\rho$ are related by $u\otimes^\rho Y =^d u(1-\rho)\otimes Y$.

Definition 2.2. A r.v. X on N_0 is called ρ-modified α-monotone((ρ,α)-MM) if it can be represented as

$$X =^d U^{1/\alpha} \otimes^\rho Y, \tag{7}$$

where Y is a r.v. defined on N_0 independent of $U \sim U(0,1)$ and $\rho \in [0,1]$.

Obviously, an $(0,\alpha)$-MM r.v. X is an α-monotone r.v. in the usual sense.

Theorem 2.1. *A r.v. X with distribution $\{p_n\}_0^\infty$ is a (ρ,α)-MM iff $X =^d \sum_{i=1}^{N} Z_i$, where $Z_1, Z_2, \ldots$ are i.i.d. Ber$(1-\rho)$ r.v.'s and N is a nonnegative integer-valued α-monotone r.v. independent of Z_i's.*

Proof. Since if X satisfies (7), by (6), we can readily obtain

$$\begin{aligned} X &=^d U^{1/\alpha} \otimes^\rho Y \\ &=^d ((1-\rho)U^{1/\alpha}) \otimes Y \\ &=^d (1-\rho) \otimes (U^{1/\alpha} \otimes Y) \\ &=^d (1-\rho) \otimes N, \end{aligned}$$

where $N =^d U^{1/\alpha} \otimes Y$ is an α-monotone r.v. and vice versa, as required. □

Corollary 2.1. *A r.v. X on N_0 is (ρ,α)-MM iff its pgf can be written as*

$$P_X(t) = \alpha[(1-\rho)(1-t)]^{-\alpha} \int_{1-(1-\rho)(1-t)}^{1} (1-w)^{\alpha-1} Q(w) d$$

where $Q(.)$ is the pgf of a r.v. on N_0.

Proof. Using Theorem 2.1, we can write

$$\begin{aligned} P_X(t) &= E(t^{(1-\rho)\otimes(U^{1/\alpha}\otimes Y)}) \\ &= P_{U^{1/\alpha}\otimes Y}(1-(1-\rho)(1-t)). \end{aligned}$$

Thus, by Theorem 1.1 and the uniqueness property of pgf's, we have the result.□

Also, it is easy to show that

Corollary 2.2. *A (ρ,α)-MM r.v. is α-monotone, but not necessarily vice versa.*

Alamatsaz [10] considered convolution property of α-monotone distributions. He proved that the convolution of a discrete α-monotone and a discrete β-monotone distribution is a discrete $\alpha+\beta$-monotone distribution. Using a similar method, we prove the following convolution property of ρ-modified α-monotone distributions.

Theorem 2.2. *Convolution of a (ρ,α)-MM and a (ρ,β)-MM distribution is a $(\rho,\alpha+\beta)$-MM distribution.*

Example 2.1. $\mathrm{Ge}_0(\pi)$ is a monotone distribution and a $\mathrm{Ge}_0(\pi)$-compounding of i.i.d. $\mathrm{Ber}(1-\rho)$ r.v.'s may be shown to be a $\mathrm{Ge}_0(p)$ distributed r.v., where $p=\frac{\pi}{1-(1-\pi)\rho}$ is a real value in $(0,1)$. Thus, by Theorem 2.1, $\mathrm{Ge}_0(p)$ is a $(\rho,1)$-MM distribution, for all $p\in(0,1)$.

3. ρ-Generalized α-Monotone Discrete Distributions

Now, we extend binomial thinning operator $\otimes$ based on inflated-parameter Bernoulli r.v.'s.

Definition 3.1. Let Y be a discrete r.v. on N_0 and $X_1, X_2, ...$ be i.i.d. $\mathrm{IBer}(u,\rho)$ r.v.'s with $\rho, u\in[0,1)$. Then, we call the operator $\otimes_\rho$ defined by

$$u\otimes_\rho Y=\sum_{i=1}^{Y} X_i \tag{8}$$

an inflated binomial thinning operator with inflating parameter ρ.

In the case $\rho=0$ (no inflation), the inflated binomial thinning operator in (8) is again the (usual) thinning operator $\otimes$ introduced in (1).

We may also note that if P_Y is the pgf of Y, for the pgf of $u \otimes_\rho Y$ we have

$$\begin{aligned} P_{u\otimes_\rho Y}(t) &= EE(t^{u\otimes_\rho Y}|Y) \\ &= EE(t^{\sum_{i=1}^{Y} X_i}|Y) \\ &= E[E(t^{X_1})]^Y \\ &= E[1-\frac{u(1-t)}{1-t\rho}]^Y \qquad \text{(from (4))} \\ &= P_Y(1-u+u\frac{t(1-\rho)}{1-t\rho}) \\ &= P_{u\otimes Y}(\frac{t(1-\rho)}{1-t\rho})\,. \end{aligned}$$

Therefore,

$$u \otimes_\rho Y =^d \sum_{i=1}^{u\otimes Y} Z_i, \tag{9}$$

where $Z_1, Z_2, ...$ are i.i.d. $\mathrm{Ge}_1(1-\rho)$ r.v.'s with pgf $P_Z(t) = \frac{t(1-\rho)}{1-t\rho}$, i.e., $u \otimes_\rho Y$ is a $u \otimes Y$-compounding of $\mathrm{Ge}_1(1-\rho)$ r.v.'s.

Remark 3.1. In the binomial thinning operator, we have $0 \otimes Y =^d 0$ and $1 \otimes Y =^d Y$, but in the inflated binomial case, we have

(a) $0 \otimes_\rho Y =^d 0$,
(b) $1 \otimes_\rho Y =^d \sum_{i=1}^{Y} Z_i$, where $Z_1, Z_2, ...$ are i.i.d. $\mathrm{Ge}_1(1-\rho)$ r.v.'s.

Now, we can define ρ-generalized α-monotone discrete distributions.

Definition 3.2. We call a r.v. X on N_0 ρ-generalized α-monotone((ρ,α)-GM) if it can be represented as

$$X =^d U^{1/\alpha} \otimes_\rho Y, \tag{10}$$

where Y is a r.v. on N_0, independent of $U \sim U(0,1)$ and $\rho \in [0,1)$ is a constant.

Obviously, a $(0,\alpha)$-GM r.v. X is α-monotone, i.e., it satisfies (2).

Theorem 3.1. *A r.v. X with distribution $\{p_n\}_0^\infty$ is (ρ,α)-GM iff $X =^d \sum_{i=1}^{N} Z_i$, where $Z_1, Z_2, ...$ are i.i.d. $Ge_1(1-\rho)$ r.v.'s, independent of nonnegative integer-valued α-monotone r.v. N.*

Proof. Since if X satisfies (10) then, by (9), we have

$$\begin{aligned} X &=^d U^{1/\alpha} \otimes_\rho Y \\ &=^d \sum_{i=1}^{U^{1/\alpha}\otimes Y} Z_i \\ &=^d \sum_{i=1}^{N} Z_i \end{aligned}$$

where $N =^d U^{1/\alpha} \otimes Y$ is an α-monotone r.v. and vice versa, as required. □

Using Theorems 1.1 and 3.1, we can easily show that

Corollary 3.1. *A r.v. X on N_0 is (ρ, α)-GM iff its pgf has the form*

$$P_X(t) = \alpha(\frac{1-t}{1-\rho t})^{-\alpha} \int_{1-\frac{1-t}{1-\rho t}}^{1} (1-w)^{\alpha-1} Q(w)dw,$$

where $Q(.)$ is the pgf of a r.v. on N_0.

Again using the method of Alamatsaz [10], we can prove analogously that the following convolution property for ρ-generalized discrete α-monotone distributions holds.

Theorem 3.2. *Convolution of a (ρ, α)-GM and a (ρ, β)-GM distribution is $(\rho, \alpha+\beta)$-GM.*

Example 3.1. Since $\mathrm{Ge}_0(\pi)$ is monotone and $\mathrm{IGe}(\pi, \rho)$ is a $\mathrm{Ge}_0(\pi)$-compounding of i.i.d. $\mathrm{Ge}_1(1-\rho)$ r.v.'s with $\rho \in [0, 1)$, it follows that $\mathrm{IGe}(\pi, \rho)$ is a $(\rho, 1)$-GM distributed r.v., for all $\pi \in (0, 1)$.

In the following theorem, we give an alternative characterization to that of Alzaid and Al-Osh [5], i.e., Theorem 1.2.

Theorem 3.3. *A discrete r.v. X on N_0 is α-monotone iff for every nonnegative bounded measurable function g and for some $\rho \in [0, 1)$, $t^\alpha E[g(t \otimes_\rho X)]$ is non-decreasing in $t \in (0, 1]$.*

Proof. Observe that for any non-negative bounded measurable g we can write

$$t^\alpha E(g(t \otimes_\rho X)) = t^\alpha E(g^*(t \otimes X)),$$

where $g^*(x) = g(\sum_{i=1}^{x} Z_i)$ belongs to the same class as g and $Z_1, Z_2, ...$ are i.i.d. $\mathrm{Ge}_1(1-\rho)$ r.v.'s. Thus, the assertion follows by Theorem 1.2. □

Remark 3.2. In Corollary 3.1, a (ρ, α)-GM r.v. was proved to be an N-compounding of i.i.d. $\text{Ge}_1(1-\rho)$ r.v.'s, with nonnegative integer-valued α-monotone r.v. N. Thus, it is reasonable to define discrete ρ-generalized α-unimodality as follows.
A r.v. X is called ρ-generalized α-unimodal ((ρ, α)-GU), if X is an N-compounding of i.i.d. $\text{Ge}_1(1-\rho)$ r.v.'s with nonnegative integer-valued α-unimodal r.v. N.
For example, Poisson distribution $\text{Po}(\lambda)$ is a unimodal r.v. and inflated-parameter distribution $\text{IPo}(\lambda, \rho)$ is a $\text{Po}(\lambda)$-compounding of i.i.d. $\text{Ge}_1(1-\rho)$ r.v.'s with $\rho \in [0, 1)$, it follows that $\text{IPo}(\lambda, \rho)$ is a $(\rho, 1)$-GU distributed r.v., for all $\lambda \in (0, \infty)$.

4. Acknowledgment

We wish to acknowledge the office of Graduate Studies of the university of Isfahan and the University of Sistan and Balochestan for their supports.

References

1. F.W. Steutel, K. van Harn, Discrete analogues of self-decomposability and stability,*Annals of Probability*, **7**(5), 893-899, (1979).
2. E. Mckenzie, Some simple models for discrete variate time series,*Water Resour. Bull.*, **21**(4), 645-650, (1985).
3. F.W. Steutel, Note on discrete α-unimodality, *Statistica Neerlandica*, **48**(2), 137-140, (1988).
4. R.A. Olshen, L.J. Savage, A generalized unimodality, *Journal of Applied Probability*, **7**(1), 21-34, (1970).
5. A.A. Alzaid, M.A. Al-Osh, Some results on discrete α-monotonicity, *Statistica Neerlandica*, **44**(1), 29-33, (1990).
6. F.W. Steutel, K. van Harn, *Infinite Divisibility of Probability of Distributions on the Real Line* (Marcel Dekker, New York, 2004).
7. C.H. Weiβ, Thinning operations for modeling time series of counts-a servey, *AStA Advances in Statistical Analysis*, Springer-verlag, **92**(3), 391-341, (2008).
8. N. Kolev, L. Minkova, P. Neytchev, Inflated-parameter family of generalized power series distributions and their application in analysis of over-dispersed insurance data, *http://www.soa.org/library/research/ actuarial-research-clearing-house/2000-09/2000/arch-2/arch00v211.pdf*, (2000).
9. L. Minkova, A generalization of the classical discrete distributions, *Communication in Statistics: Theory and Methods*, **31**(6), 871-888, (2002).
10. M.H. Alamatsaz, On discrete α-unimodal distributions, *Statistica Neerlandica*, **47**(4), 245-252, (1993).

Application And Enhancement Of Monte Carlo Integration

M. Alijani and A. Arshadipour

Department of Statistics, Razi University
Department of Mathematical Sciences, Isfahan University of Technology

Randomness of Monte Carlo approximation for the access of integrals, specially in fractional calculus, reduces the precision of them with comparison to deterministic numerical methods. In this paper, we enhance this approximation by the utilization of the Riemann method just in uniform cases. After that we compare traditional Monte Carlo by an example and particular attention is devoted to the modified Monte Carlo integrals in order to simulate fractional calculus.

keywords: Monte Carlo; Riemann approximation; Random numbers; Fractional derivatives; Fractional integrals.

1. Introduction

In complex cases, bounded integrals are approximated by using relevant methods, for example Riemann and Monte Carlo methods (Daivis and Rabinwitz (1984)). Although, numerical methods are approximated of them but, the answer of integrals have some differences. In statistical approximation Monte Carlo methods is based on independent random samples from statistical distribution. Randomness factor causes different answers in each runs. On the other hand, Fractional derivatives and integrals in the framework of the Riemann-Liouville fractional calculus have complicated forms that could be achieved by Tailor extension,(Oldham (1974), Miller and Ross (1993)) but complexity in calculation of them causes not to find theirs applications. We compare it by an example in which the Riemann's numerical answer for bounded integrals is more scholastic than Monte Carlo integrals in uniform cases. It should be noted that there are two issues in Monte Carlo integrations. Randomness and knowledge of the distribution. The bounded integrals is in form of $I = \int_a^b f(x)dx$ and we can use random samples in uniform $[a, b]$ and evaluate the integral by Monte Carlo method in this case. We use novel techniques of simulations to accomplish them and depict the related figures. We prefer to reintroduce the fractional derivatives, integrals and Monte Carlo methods in this part. Then approximation of fractional derivations and integrals by simulations for different quantities are discussed, and we do simulation 1000 times

for any quantity, after that we will draw their figures. The function $f(x) = sin(x^2)$ is as the simplest example in this paper and the approximate fractional derivative and integral. It is easy to do this, simulation method for more complex functions, that it is hard to integrate them, with hope to find more applications of fractional calculus in the future. Deterministic and stochastic methods for evaluation of integrals are discussed in this section. In section two, we will compare deterministic and stochastic methods with enhance stochastic Monte Carlo methods. Section three consists of fractional calculus and simulation of them is the forth section of this paper.

2. Deterministic Methods to Evaluation of Integrations

evaluation of the integral $I = \int_a^b f(x)dx$ is a problem as old as the calculus itself and is equivalent to solution of the differential equation $\frac{dy}{dx} = f(x)$ subject to the boundary condition $y(a) = 0$. In well-catalogued instances, analytical solutions are available. This means neglecting the classical but dated approaches using equally spaced abscissas, and a more extended discussion may be found in Davis and Rabinowitz (1984, Chapter 2). The principle underlying most state-of-the-art deterministic evaluations of $I = \int_a^b f(x)dx$ is Riemann numerical method. In this method, we have:

$$\int_a^b f(x)dx \cong \sum_{i=1}^{n} f(x_i)\Delta(x_i) \qquad \Delta(x_i) = x_{i+1} - x_i \qquad x_i \in [a, b] \tag{1}$$

n is number of x_i in $[a, b]$.

2.1. *Monte Carlo Integrations*

Perhaps the earliest documented use of random sampling to find the solution to an integral is that of Comte de Buffon. In 1777 he described a needle experiment (Christian and Casella (2004)). Monte Carlo simulation is the use of experiment with random number to evaluate mathematical expressions (Alijani and Drikvandi (2009)). The experimental units are the random numbers and the expressions may be definite integrals, system of equation, or more complicated mathematical models. In most cases, when a mathematical expression is to be evaluated, the standard approximation from numerical analysis is to be preferred, but Monte Carlo methods provide an alternative that is sometimes only a tractable approach. Monte Carlo is often the preferred method for evaluating integrals over high-dimensional domains (Alijani and Ghoreishi (2009)). Very large and sparse systems of equation can sometimes be solved effectively by Monte Carlo methods. In its simplest form,

Monte Carlo simulation is the evaluation of integrals:

$$\Theta = \int_a^b f(x)dx \tag{2}$$

by identifying from a random variable Y with support on $[a, b]$ and density $p(y)$ and a function g so that the expected value of $g(Y)$ is Θ :

$$E(g(Y)) = \int_a^b g(y)p(y)dy = \int_a^b f(y)dy = \Theta. \tag{3}$$

Let us first consider the case in which $[a, b]$ the integral Y is taken to be a random variable with a uniform density over $[a, b]$, and g is taken to be f. In this case,

$$\Theta = (b - a)E(f(Y)). \tag{4}$$

The problem of evaluating the integral becomes the familiar statistical problem of estimating the integral, a mean $E(f(Y))$. The statistician quite naturally takes a random sample and use the sample mean. For a sample of size m, an estimate of Θ is

$$\hat{\Theta} = (b - a)\frac{\sum_{i=1}^m f(y_i)}{m} \tag{5}$$

where the y_i is value of a random sample from uniform distribution over $[a, b]$. This estimator is unbiased. Because:

$$E(\hat{\Theta}) = (b - a)\frac{E(f(Y_i))}{m} = (b - a)E(f(Y)) = \int_a^b f(x)dx = \Theta. \tag{6}$$

3. Comparison Between Deterministic Riemann and Stocastic Monte Carlo Methods

In Monte Carlo method, if X_i is selected in uniform $[a, b]$, after comparison with Reimann methods, so:

$$\int_a^b f(x)dx \cong \frac{(b - a)}{n}\sum_{i=1}^n f(x_i) \tag{7}$$

$$\int_a^b f(x)dx \cong \sum_{i=1}^n f(x_i)\Delta(x_i) \tag{8}$$

In deterministic Reimann method x_i's isn't random sample. We can reduce the stochastic factor of Monte Carlo method by deterministic Riemann method. After that, we have:

$$\int_a^b f(x)dx \cong \frac{(b - a)}{n}\sum_{i=1}^n f(x_i) \cong \sum_{i=1}^n f(x_i)\Delta(x_i) \tag{9}$$

If we change $\frac{(b-a)}{n}$ in middle part of (9) by $\Delta(x_i)$ and sort the random samples of X_i then calculate $\Delta(x_i)$ and set the Monte carlo approximation, the value of this estimator has more precision and is nearer to the real value of integral. For distinction, in this paper, we nominate the traditional Monte Carlo by M_t and sorted Monte Carlo by M_s. The variance of Monte Carlo estimator has important uses in assessing the quality of the integrals, where the X_i is value of a random sample from uniform distribution over $[a, b]$. M_s is asymptotic unbiased and the variance of it is less than variance of M_t. We proved two new theorems about this facts.

Theorem 1.

If $I = \int_a^b f(x)dx$, $\hat{I} = \sum_{i=1}^n f(x_i)\Delta x_i$ and $X_i^* \sim Uniform[a, b]$, $x_i = sort(X_i^*)$ then

$E(M_s) \to (b-a)E(f(x)) = I$

Proof

$$\begin{aligned} E(M_s) &= E(\sum_{i=1}^{n} f(X_i)\Delta X_i) = \sum_{i=1}^{n} E(f(X_i))E(\Delta X_i) \\ &= \frac{b-a}{n+1}\sum_{i=1}^{n} E(f(X_i)) = \frac{b-a}{n+1}E(\sum_{i=1}^{n}(f(X_i))) = \frac{n(b-a)}{n+1}E(f(X)) \\ &\Rightarrow \lim_{n\to\infty} \frac{n(b-a)}{n+1}E(f(X)) = (b-a)E(f(X)) = I. \end{aligned}$$

Theorem 2.

If $I = \int_a^b f(x)dx$, $\hat{I} = \sum_{i=1}^n f(x_i)\Delta x_i$ and $X_i^* \sim Uniform[a, b]$, $x_i = sort(X_i^*)$ then

$Var(M_s) \leq Var(M_t)$.

Proof

$$\begin{aligned} Var(M_s) =& Var(\sum_{i=1}^{n} f(x_i)\Delta(x_i)) \\ =& \sum_{i=1}^{n} Var(f(x_i)\Delta(x_i)) + \sum_{i\neq j} Cov(f(x_i)\Delta(x_i), f(x_j)\Delta(x_j)) \\ =& (b-a)^2\{\sum_{i=1}^{n} E(f^2(X_i))\frac{n}{(n+1)^2(n+2)} - (E(f(X_i)))^2\frac{1}{(n+1)^2}\} \\ &+ \frac{1}{(n+1)^2}\sum_{i\neq j} Cov(f(x_i), f(x_j)) \end{aligned}$$

$$
\begin{aligned}
&\leq (b-a)^2\{\sum_{i=1}^{n}\frac{n}{(n+1)^2(n+2)}(E(f^2(X_i))) - (E(f(X_i)))^2\} \\
&\quad + \frac{1}{(n+1)^2}\sum_{i\neq j} Cov(f(x_i), f(x_j)) \\
&\leq (b-a)^2\left(\frac{1}{(n+1)^2}(\sum_{i=1}^{n} Varf(x_i)) + \sum_{i\neq j} Cov(f(x_i), f(x_j))\right) \\
&\leq (b-a)^2\left(\frac{1}{(n+1)^2} Var(\sum_{i=1}^{n} f(x_i))\right) \\
&\leq \frac{(b-a)^2}{n^2} Var(\sum_{i=1}^{n} f(x_i)) = Var(M_t).
\end{aligned}
$$

Example 1.

In simplest case, we are ready to calculate $\int_{-5}^{5} sin(x^2)dx$ by two methods. Before computations, in figure 1 this function in interval $[-5, 5]$ is depicted and we find out integration of $sin(x^2)$ is positive in mentioned interval. Real value of its is equal to $\int_0^{\frac{5\sqrt{2}}{\sqrt{\pi}}} \sqrt{2\pi} cos(\frac{\pi}{2t^2})dt$. By separating interval of $[-5, 5]$ to 10000 equal partition and calculation of denoted integral by simple numerical method, we have:

$$\int_{-5}^{5} sin(x^2)dx \cong \sum_{i=1}^{9999} sin(x_i^2)\Delta(x_i) = 0.000142256, \quad \Delta(x_i) = x_{i+1} - x_i$$

On the other hand, if we use Monte Carlo methods and calculate this integral by 10000 random samples from uniform $[-5, 5]$ we have:

$$\int_{-5}^{5} sin(x^2)dx \cong \frac{(b-a)}{n}\sum_{i=1}^{10000} f(x_i) = \frac{1}{1000}\sum_{i=1}^{10000} sin(x_i^2) = 0.0003840335$$

it means that, $\hat{\mu}_1 = M_t = 0.0002840335$

We can find the variance of Monte Carlo approximation by 10(Christian and Casella (2004)).

$$Var(\hat{\mu}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(f(x_i) - \hat{\mu})^2 \tag{10}$$

After calculation for variance of this quantity equals to $Var(\hat{\mu}_1) = Var(M_t) = 5.472017e-09$

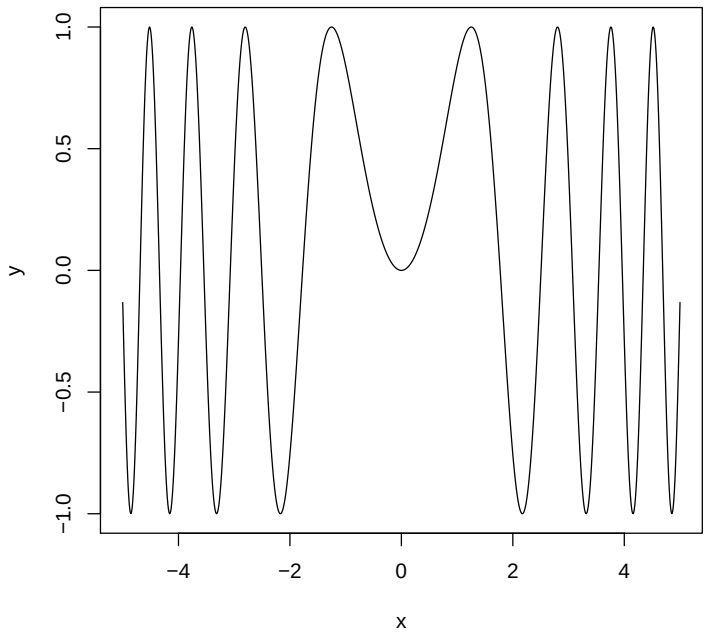

Fig. 1. $f(x) = sin(x^2)$

For computation of $\int_{-5}^{5} sin(x^2)dx$ by Modified Monte Carlo method (M_s), the random samples should be sorted calculated $\sum_{i=1}^{9999} sin(x_i^2)\Delta(x_i)$ in that $\Delta x_i = x_{i+1} - x_i$. Then,

$$\int_{-5}^{5} sin(x^2)dx \cong \sum_{i=1}^{9999} sin(x_i^2)\Delta(x_i) = 0.0002513409.$$

After calculation for variance of this quantity equal to $Var(\hat{\mu}_2) = Var(M_s) = 1.472017e{-}09$ variance in this case is less than traditional Monte Carlo method. We have demonstrated a simple way to improve the calculation of bounded integrals by using deterministic Riemann methods, and depicting their result by example. Indeed, reduction of the stochastic factor helps us to decrease its variance and this approximation is being nearer to the real value. In cases of Monte Carlo we use with uniform density between two bounds of integrals and, this method can be used to evaluated the complex function.

4. Fractional Calculus

According to the Riemann-Liouville approach to fractional calculus the notion of fractional integral of order $\alpha(\alpha > 0)$ is a natural consequence of the well known formula(uaually attributed to Caushy),that reduces the calculation of the n-fold primitive of function $f(t)$ to a single integral of convolution type (Miller and Ross(1993)). In our notation the Caushy formula reads

$$J^n f(t) := f_n(t) = \frac{1}{(n-1)!}\int_0^t (t-\tau)^{n-1} f(\tau)d\tau \quad , t > 0 \quad , n \in N \tag{11}$$

where N is positive integer. From this definition, it is noted that $f_n(t)$ vanishes at $t = 0$ with its derivatives of order $1, 2, \cdots, n-1$. For convention we require that $f(t)$ and henceforth $f_n(t)$ to be a causal function, i.e identically vanishing for $t < 0$. In a natural way one is led to extend the above formula from positive integer values of the index to any positive real values by using the Gamma function. Indeed, nothing that $(n-1)! = \Gamma(n)$, and introducing the arbitrary positive real number α, one define the Fractional Integral of order α :.

$$J^\alpha f(t) := \frac{1}{\Gamma(\alpha)}\int_0^t (t-\tau)^{\alpha-1} f(\tau)d\tau \quad , t > 0 \quad , \alpha \in R^+ \tag{12}$$

where R^+ is the set of positive real numbers . For complementation is defined $J^0 := I$ (Identity operator),i.e. $J^0 f(t) = f(t)$. Furthermore, by $J^\alpha f(0^+)$ we mean the limit (if it exists) of $J^\alpha f(t)$ for $t \to 0^+$; this limitation may be infinite.
After the notation of fractional integral, that of fractional derivative of order α $(\alpha > 0)$ becomes a natural requirement and one is attempted to substitute α with $-\alpha$ in the 12. However, this generalization needs some care in order to guarantee the convergence of the integrals and preserves the well known properties of the ordinary derivative of integer order. Denoting by D^n with $n \in N$, the operator of the derivative of order n, we first note that

$$D^n J^n = I, \quad J^n D^n \neq I, \quad n \in N \tag{13}$$

$$J^n D^n f(t) = f(t) - \sum_{k=0}^{n-1} f^{(k)}(0^+)\frac{t^k}{k!}, \quad t > 0. \tag{14}$$

As a consequence we expect that D^α is defined as left-inverse to J^α. For this purpose, introducing the positive integer m such as $m-1 < \alpha \leq m$, one define the Fractional Derivative of order $\alpha > 0 : D^\alpha f(t) := D^m J^{M-\alpha} f(t)$ namely (Miller and Ross (1993)).

$$D^\alpha f(t) := \frac{d^m}{dt^m}[\frac{1}{\Gamma(m-\alpha)}\int_0^t \frac{f(\tau)}{(t-\tau)^{\alpha+1-m}}d\tau] \quad m-1 < \alpha < m. \tag{15}$$

and for $\alpha = m$ we have $D^{\alpha} f(t) := \frac{d^m}{dt^m} f(t)$ Defining for complementations $D^0 = J^0 = I$ then we easily recognize that

$$D^{\alpha} J^{\alpha} = I, \quad \alpha \geq 0 \tag{16}$$

and

$$D^{\alpha} t^{\gamma} = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1-\alpha)} t^{\gamma-\alpha}, \quad \alpha > 0 \;\; \gamma > -1 \;\; t > 0. \tag{17}$$

Note the remarkable fact that the fractional derivative $D^{\alpha} f$ is not zero for the constant function $f(t) = 1$ if $\alpha \in N$.In fact 17 with $\gamma = 0$ teaches us that

$$D^{\alpha} 1 = \frac{t^{-\alpha}}{\Gamma(1-\alpha)}, \;\; \alpha \geq 0 \;\; t > 0. \tag{18}$$

this of course is$\equiv$ 0 for $\alpha \in N$ due to poles of *Gamma* function in the point $0, -1, -2, \cdots$. We now observe that an alternative definition of fractional derivative originally introduce by Caputo and Mainardi (1971) in the frame work of the theory of Linear Viscoelasticity (see Baleanu and Avkar (2004)), is so-called Caputo Fractional Derivative of order $\alpha > 0 : D_*^{\alpha} f(t) := J^{m-\alpha} D^m f(t)$ with $m - 1 < \alpha \leq m$, and this formula for $m - 1 < \alpha < m$ is

$$\frac{1}{\Gamma(m-\alpha)} \int_0^t \frac{f^{(m)}(\tau)}{(t-\tau)^{\alpha+1-m}} d\tau$$

then, for $\alpha = m$ is

$$\frac{d^m}{dt^m} f(t)$$

5. Simulation of Eractional Calculus by Monte Carlo Methods

A fractional derivative of order α is given by using Caputo definition for $m - 1 < \alpha < m$ is

$$\frac{1}{\Gamma(m-\alpha)} \int_0^t \frac{f^{(m)}(\tau)}{(t-\tau)^{\alpha+1-m}} d\tau$$

We can simulate this form of integrals by using modified Monte Carlo methods.
Algorithm for generation of $D^{\alpha} f(x(t_1)), \cdots, D^{\alpha} f(x(t_n))$
$1 - Let \;\; y_{(i)} = generate \;\; m \;\; random \;\; number \;\; from \;\; uniform \;\; (0, t_i) \quad j = 1, \cdots, m, \;\; i = 1 \ldots, n \;\; and \;\; sorted \;\; (y)$
$2 - Let \; z_j = \left(f^k(y_j) / (t_i - y_j)^{\alpha+1-k} \right) \times \Delta(y_j), \quad k - 1 < \alpha < k$
$3 - Let \; D^{\alpha}(f(x(t_i)) = \frac{1}{\Gamma(k-\alpha)} \times \sum z_j$
For example we have simulated $D^{\alpha}(sin(x^2))$ for $\alpha = 0.1, \alpha = 0.2, \cdots, \alpha = 0.9$
And this problem is solved in R software, after that it is shown in figure 2. When α is increased, a particular trend will be found.

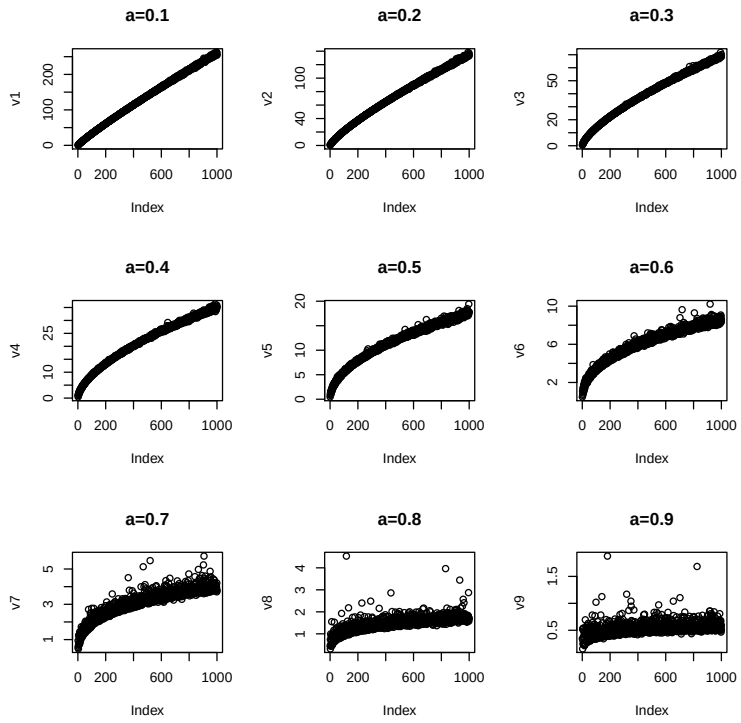

Fig. 2. Simulation from $D^{\alpha}(sin(x^2))$ for $\alpha = 0.1, \alpha = 0.2, \cdots, \alpha = 0.9$.

We prefer to write again, Rieman-Liouville definition of fractional integral of order $\alpha > 0$, which is given as [8]

$$I^{\alpha}y(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} f(\tau) d\tau, \quad (\alpha > 0) \tag{19}$$

and we can do this form of integral by using modified monte carlo method.
Algorithm for generation samples from $I^{\alpha}(y(t_1)), I^{\alpha}f(y(t_2)), \cdots, I^{\alpha}f(y(t_n))$:
$1 - Let\ y_j = generate\ m\ random\ number\ from\ uniform\ (0, t_i)\ j = 1, 2, \cdots, m,\ \ i = 1 \ldots, n\ \ and\ \ sorted\ (y)$
$2 - Let\ z_j = \big((t_i - y_j)^{\alpha-1}/(f(y_j)\big) \times \Delta(y_j)$
$3 - Let\ D^{\alpha}(f(x(t_i)) = \frac{1}{\Gamma(\alpha)} \times \sum z_j$

For example, $I^{\alpha}sin(x^2)$ for $\alpha = 0.1, \alpha = 0.2, \cdots, \alpha = 0.9$, is simulated and solved in R software which is shown in figure 3. When α is increased, despite of trend shown in figure 2 a reverse trend will be found in figure 3.

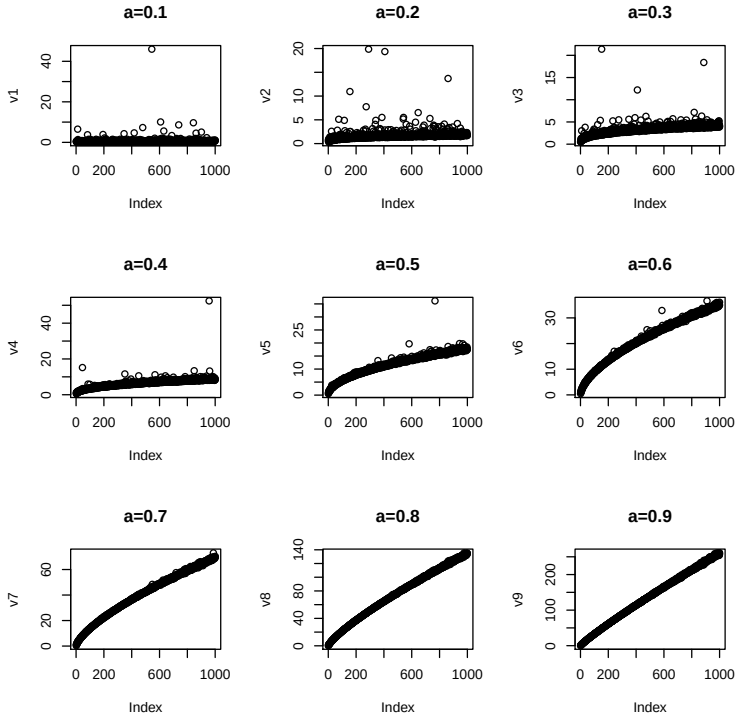

Fig. 3. Simulation from $I^{\alpha}(sin(x^2))$ for $\alpha = 0.1, \alpha = 0.2, \cdots, \alpha = 0.9$.

6. Conclusion

We have demonstrated a simple way of improving for computations of bounded integrals, using deterministic Riemann methods, and depicting their result by example. Indeed, reduction of the stochastic factor help us to decrease of its variance and this approximation is being nearer to real value. In cases of Monte Carlo using with uniform density between two bounds of integrals, we can use this method for

complicated computations of fractional derivations and integrals, using approximate Monte Carlo methods, and depicted their figures for a variety of amounts α. Some of these points have not been exactly fitted on relevant carve due to their stochastic aspects. But in general, current trend in these figures have been considered to create their novel applications in the field of physics (Samko et al (1993), Miller and Ross (1993), Hilfer (2000)) and the other sciences (Ghoreishi et al (2009), Mainardi and Carpinteri (1997)).

References

1. Alijani, M. and Ghoreishi, S.K., *Simulation of fractional integrals and derivatives by stochastic methods* , Proceeding of 2th Mathematical Science Conference of PNU, (2009).
2. Alijani, M. and Drikvandi, G.R., *Modified Monte Carlo Integrals In Uniform Cases.* , Proceeding of 2th Mathematical Science Conference of PNU, (2009).
3. Baleanu, D. and Avkar, T., *Lagrangians with linear velocities within Riemann-Liouville fractional derivatives.*, Nuovo Cimento Soc. Ital., B, 119(2004)73-79.
4. Caputo, M. and Mainardi, F., *Linear models of dissipation in anelastic solids.*, Rivista. Nuovo Cimento, 1(1971)161-198.
5. Davis, P. J. and Rabinowitz, P., *Methods of Numerical Integration.*, Orlando, Academic press. (second edition)., (1984).
6. Christian, P. R. and Casella, G., *Monte Carlo Statistical Methods.*, Springer-Verlag, New York., (2004).
7. Ghoreishi, S. K. and Alijani, M. and Yaghmaei, F., *Simulation of Fractional Calculus by Monte Carlo Methods.*, New York: Academic Press. *Proceeding of SPSP7.*, (2009).
8. Hilfer. R., *Applications of Fractional Calculus in Physics.*, World Scientific, New Jersey., (2000).
9. Mainardi, F. and Carpinteri, A., *Fractals and Fractional Calculus in Continuum Mechanics.*, Springer-Verlag, New York, (1997).
10. Miller, S. and Ross, B. , *An Introduction to the Fractional Calculus and Fractional Differential Equations.*, John Wiley and Sons., (1993).
11. Oldham, K.B., *The Fractional Calculus.*, Academic Press, New York., (1974).

On Unimodality of Generalized Order Statistics and Their Dual

M. Alimohammadi and M. H. Alamatsaz

Statistics Department, University of Isfahan,

Generalized order statistics (GOSs) and their dual (DGOSs) have been introduced as a unification of several models of random variables arranged in ascending and descending order of magnitude with different interpretations and statistical applications, respectively. Some authors have worked on strong unimodality of GOSs under some rather strong conditions. In this article, we shall concentrate to the unimodality of these statistics in general. To do this, we shall first correct an erroneous result of Basak and Basak (2002) in establishing the unimodality of record values and generalize this result to k-record values. Then, we shall extend the result, more generally, to certain important cases of GOSs and DGOSs which contain Alam's (1972) results as corollaries.

Keywords: Unimodality; Order statistics; Upper k-record values; Lower k-record values; Generalized order statistics; Dual generalized order statistics; Convexity.

1. Introduction

A distribution function $F(x)$ is said to be unimodal if there exists a value $x = a$ such that $F(x)$ is convex for $x < a$ and concave for $x > a$ and $F(x)$ is said to be strongly unimodal if its convolution with any unimodal distribution is unimodal. Dharmadhikari and Joag-Dev (1988) is a good monograph for these concepts. Huang and Ghosh (1982) and Chen et al. (2009) have worked on strong unimodality of ordinary order statistics and GOSs, respectively, under some rather strong conditions. For a distribution function $F(x)$ with density $f(x)$, Alam (1972) showed that the condition of convexity of $\frac{1}{f(x)}$ is sufficient to ensure unimodality of the corresponding ordinary order statistics. Basak and Basak (2002) investigated unimodality of record values, but Aliev (2003) provided a counter example showing that the convexity of $\frac{1}{f(x)}$, formulated by Basak and Basak, is not sufficient to conclude that the upper record values are unimodal. In this paper, we shall first modify Basak and Basak's result and establish the result for unimodality of k-record values. Then, we shall extend the result, more generally, to certain important cases of GOSs and DGOSs which contain Alam's results.

2. Unimodality of record statistics

Let $X_1, X_2, ...$ be iid random variables (r.v.) with absolutely continuous cumulative distribution function (cdf) F and probability density function (pdf) f. For a sample size of $n \geq 1$, we denote the order statistics of $X_1, X_2, ..., X_n$ by $X_{1,n}, X_{2,n}, ..., X_{n,n}$. Then the sequence $\{T_n^{(k)}, n \in N\}$, $k \in N$, with nth upper k-record instances

$$\begin{aligned} T_1^{(k)} &= 1, \\ T_{n+1}^{(k)} &= min\{j \in N : j > T_n^{(k)}, X_{j,j+k-1} > X_{T_n^{(k)}, T_n^{(k)}+k-1}\}, \quad n \in N \end{aligned} \tag{1}$$

and nth upper k-record values

$$U_n^{(k)} = X_{T_n^{(k)}, T_n^{(k)}+k-1}, \quad n \in N \tag{2}$$

are called nth upper k-record statistics. Similar definition is given for the nth lower k-record statistics $L_n^{(k)}$ by switching the last inequality in Eq. (1). Obviously, we obtain ordinary upper record values when $k = 1$ and, in this case, we denote $U_n^{(1)}$ by $X(n)$. For more details on record statistics, we refer to Arnold et al. (1998) and Nevzorov (2001).

Basak and Basak claimed the following incorrect results for unimodality of upper record values. First, we shall reveal the incorrectness of their proof, then attempt to modify their theorems to show the unimodality of k-record values, in general.

Theorem 2.1 (Basak and Basak, 2002). *(i) Suppose that the pdf f of cdf F is such that $1/f$ is convex. Then, each upper record statistic $X(n)$, $n \geq 1$, has a unimodal distribution.*

(ii) Let the cdf F be unimodal with mode at a, and the pdf f. Let f be continuous at a, and $1/f$ be convex in $x > a$. Then, $X(n)$, $n \geq 1$ is unimodal.

Proof. (i) The density function $f_n(x)$ of $X(n)$ is given by

$$f_n(x) = \frac{1}{(n-1)!}(H(x))^{n-1} f(x), \quad x \in R \tag{3}$$

where $\bar{F}$ and $H(x) = -ln\bar{F}(x)$ are the survival function and the cumulative hazard rate of X, respectively (see, e.g., Arnold et al. (1998), p. 10). If we denote the hazard rate of X by $H'(x) = h(x) = f(x)/\bar{F}(x)$, by differentiation we obtain

$$\begin{aligned} f_n'(x) &= \frac{1}{(n-1)!}\Big[(n-1)h(x)(H(x))^{n-2} f(x) + (H(x))^{n-1} f'(x)\Big] \\ &= \frac{(H(x))^{n-1} h(x) f(x)}{(n-1)!}\Big[\frac{(n-1)}{H(x)} + \frac{f'(x)}{f^2(x)}\bar{F}(x)\Big]. \end{aligned} \tag{4}$$

But convexity of $1/f$ implies that f'/f^2 is non-increasing. Also, clearly H is non-decreasing. Basak and Basak stated that the product of two non-increasing functions f'/f^2 and $\bar{F}$ is non-increasing, then, by (4), concluded that $f_n(x)$ is unimodal.

This claim may not be true when f'/f^2 is negative. Since $\bar{F}$ is non-negative, one way to tackle this problem is to restrict ourselves to distributions for which f' is non-negative. This is, obviously, the case when f is non-decreasing on its support. However, it should be noted that since the pdf (3) of the upper record statistic, in this special case, becomes the product of two non-decreasing functions, the assumption of $1/f$ being convex is not required. Thus, this kind of distributions can have unimodal upper records (at the end point of their support).

(ii) It may be noted from (i) that, distributions for which f is unimodal and continuous at its mode can not have unimodal upper record values. Thus, this part of the theorem is not true at all, confirming Aliev (2003). □

Therefore, we arrive at the following modified and also generalized version of Theorem 2.1 as follows:

Theorem 2.2. *Let the pdf f of cdf F be such that for $k = 1$, f is non-decreasing on its support and for $k = 2, 3, ..,$ both $1/f$ is convex and f is non-decreasing on its support. Then, each upper k-record statistic $U_n^{(k)}$, $(k, n \in N)$, has a unimodal distribution.*

Proof. The density function $f_{U_n^{(k)}}(x)$ of $U_n^{(k)}$ is given by

$$f_{U_n^{(k)}}(x) = \frac{k^n}{(n-1)!}(H(x))^{n-1}(\bar{F}(x))^{k-1}f(x), \quad x \in R \tag{5}$$

(see, e.g., Arnold et al. (1998), p. 81).
So, we have

$$\begin{aligned} f'_{U_n^{(k)}}(x) &= \frac{k^n}{(n-1)!}\Big[-(k-1)f^2(x)(H(x))^{n-1}(\bar{F}(x))^{k-2} \\ &\quad +(n-1)h(x)(H(x))^{n-2}(\bar{F}(x))^{k-1}f(x) + f'(x)(H(x))^{n-1}(\bar{F}(x))^{k-1}\Big] \\ &= \frac{k^n}{(n-1)!}(H(x))^{n-1}(\bar{F}(x))^{k-1}f(x)h(x) \\ &\quad \times\Big[-(k-1) + \frac{(n-1)}{H(x)} + \frac{f'(x)}{f^2(x)}\bar{F}(x)\Big]. \end{aligned} \tag{6}$$

Thus, by the above argument, the quantity inside the brackets on the right-hand side of Eq. (6) is non-increasing in x and the quantity outside the brackets is positive. Hence, $f'_n(x)$ has at most one change of sign as x moves from $-\infty$ to ∞, from positive to negative. Therefore, $f_{U_n^{(k)}}(x)$ is unimodal. □

Example 2.1. Consider the right-truncated Normal distribution, $N(\mu, \sigma^2)$, at μ. Practically, we consider the case when $\mu > 0$ is far from zero, e.g., $\mu > 4\sigma$. In this case, the corresponding r.v. takes positive values which is desired in such contents. Obviously, this distribution satisfies the conditions of Theorem 2.2, thus, all of its upper k-record values are unimodal. Therefore, it can be concluded that distributions satisfying $1/f$ being convex, can be truncated as desired.

As other examples, consider generalized Pareto distribution with the shape parameter $c > 1$:

$$f(x) = a^{-1}(1 - cx/a)^{c^{-1}-1}, \quad 0 < x < a/c.$$

This satisfies the condition of Theorem 2.2 for $k = 1$, because it is only non-decreasing (see Johnson et al. (1994), p. 615). Particular Beta, e.g., power function distribution

$$f(x) = px^{p-1}, \quad 0 \leq x \leq 1$$

with parameter $p \geq 1$, satisfies the conditions for $k \in N$, but,

$$f(x) = q(1 - x)^{q-1}, \quad 0 \leq x \leq 1$$

where $0 < q < 1$, does not satisfy the condition of $1/f$ being convex (see Johnson et al. (1995), p. 219-220).

Next theorem shows that for unimodality of the lower case, the condition of $1/f$ being convex for any $k \in N$ is sufficient.

Theorem 2.3. *Let the pdf f of cdf F be such that $1/f$ is convex. Then, each lower k-record statistic $L_n^{(k)}$, $(k, n \in N)$, has a unimodal distribution.*

Proof. The density function $f_{L_n^{(k)}}(x)$ of $L_n^{(k)}$ is given by

$$f_{L_n^{(k)}}(x) = \frac{k^n}{(n-1)!}(-lnF(x))^{n-1}(F(x))^{k-1}f(x). \tag{7}$$

So, we have

$$\begin{aligned}
&f'_{L_n^{(k)}}(x) \\
&= \frac{k^n}{(n-1)!}\Big[(k-1)f^2(x)(-lnF(x))^{n-1}(F(x))^{k-2} \\
&\quad -(n-1)\frac{f^2(x)}{F(x)}(-lnF(x))^{n-2}(F(x))^{k-1} + f'(x)(-lnF(x))^{n-1}(F(x))^{k-1}(x)\Big] \\
&= \frac{k^n}{(n-1)!}(-lnF(x))^{n-1}(F(x))^{k-1}(x)f^2(x)\Big[\frac{(k-1)}{F(x)} + \frac{(n-1)}{lnF(x)}\frac{1}{F(x)} + \frac{f'(x)}{f^2(x)}\Big].
\end{aligned} \tag{8}$$

Hence, the assertion follows similar to that of Theorem 2.2. □

Further, in the special cases we have:

Theorem 2.4. *Suppose that the cdf F with pdf f is unimodal with mode at M. Let f be continuous at M, and $1/f$ be convex in $x < M$ $(x > M)$. Then, $L_n^{(1)}$ $(L_1^{(k)})$, $(k, n \in N)$, is unimodal.*

Proof. From (7), the pdf of $L_n^{(1)}$, the lower record statistic, is given by

$$f_{L_n^{(1)}}(x) = \frac{1}{(n-1)!}(-lnF(x))^{n-1} f(x). \tag{9}$$

Clearly, $f_{L_n^{(1)}}(x)$ is non-increasing for $x \geq M$. Since $1/f$ is convex, using the same argument as in the Theorem 2.3, $f_{L_n^{(1)}}(x)$ is either non-increasing or it is first increasing and then decreasing for $x < M$. Therefore, $f_{L_n^{(1)}}(x)$ is unimodal. For $L_1^{(k)}$, the proof is similar in reverse direction, i.e., for $x > M$. □

3. Unimodality of generalized order statistics and their dual

Generalized order statistics (GOSs) were introduced by Kamps (1995a,b) as a unification of several models of random variables arranged in ascending order of magnitude with different interpretations and statistical applications. In other words, he imposed several parameters into a model so that, by changing them, it could cover previous known ordered random variables arranged in ascending order of magnitude. Then, Burkschat et al. (2003) proposed a dual model that enables a common approach to descending ordered random variables. In this section, we investigate the unimodality of GOSs and DGOSs.

Let $n \in N$, $k \geq 1$, $m_1, m_2, ..., m_{n-1} \in R$, $M_r = \sum_{j=r}^{n-1} m_j$, $(r = 1, ..., n-1)$, be such that $\gamma_r = k + n - r + M_r \geq 1$ for all $r \in \{1, ..., n-1\}$, and let $\tilde{m} = (m_1, m_2, ..., m_{n-1})$, if $n \geq 2$ ($\tilde{m} \in R$ arbitrary, if $n = 1$).

If $m_1 = m_2 = ... = m_{n-1} = m$, we denote GOSs (DGOSs) by $X(r, n, m, k)$ $(X_d(r, n, m, k))$, $1 \leq r \leq n$. In many cases, this assumption of equality of $\tilde{m}$ components naturally holds. This assumption is not far from reality because it is satisfied by most of the well known models such as, ordinary order statistics, record values and in general k-record values, order statistics with non-integral sample size, some cases of sequential order statistics and some cases of Pfeifer's record model. Therefore, we may well assume that $m_i = m$, $i = 1, ..., n-1$.

Now, let $c_{r-1} = \prod_{i=1}^{r} \gamma_j$, $r = 1, ..., n$ with $M_n = 0$ and on the unit interval define $g_m(x)$, $(m \in R)$, by

$$g_m(x) = \begin{cases} \frac{1}{m+1}(1-(1-x)^{m+1}) & ,\text{m} \neq -1 \\ -ln(1-x) & ,\text{m} = -1 \end{cases}, \quad x \in [0, 1). \tag{10}$$

In such a case, the pdf of rth GOS and DGOS of an absolutely continuous distribution function F with pdf f are, respectively, given by

$$f_{X(r,n,m,k)}(x) = \frac{c_{r-1}}{(r-1)!}(\bar{F}(x))^{\gamma_r-1}(g_m(F(x)))^{r-1}f(x), \quad x \in R \tag{11}$$

(see Kamps (1995a), p. 64), and

$$f_{X_d(r,n,m,k)}(x) = \frac{c_{r-1}}{(r-1)!}(F(x))^{\gamma_r-1}(g_m(\bar{F}(x)))^{r-1}f(x), \quad x \in R. \tag{12}$$

Theorem 3.1. *Let F be the cdf with pdf f. Then, its GOS is unimodal if one of the following conditions are satisfied:*
(i) $m \neq -1$, $\frac{1}{f}$ is convex,
(ii) $m = -1$:
(a) $k = 1$ and f is non-decreasing on this support,
(b) $k > 1$, $\frac{1}{f}$ is convex and f is non-decreasing on its support.

Proof. (i) From (11) we have

$$\begin{aligned} f'_{X(r,n,m,k)}(x) &= \frac{c_{r-1}}{(r-1)!}\Big[-(\gamma_r-1)f^2(x)(\bar{F}(x))^{\gamma_r-2}(g_m(F(x)))^{r-1} \\ &\quad +(r-1)f^2(x)(\bar{F}(x))^m(\bar{F}(x))^{\gamma_r-1}(g_m(F(x)))^{r-2} \\ &\quad +f'(x)(\bar{F}(x))^{\gamma_r-1}(g_m(F(x)))^{r-1}\Big] \\ &= \frac{c_{r-1}}{(r-1)!}f^2(x)(g_m(F(x)))^{r-1}(\bar{F}(x))^{\gamma_r-1} \\ &\quad \times\Big[-\frac{(\gamma_r-1)}{\bar{F}(x)}+\frac{(r-1)}{g_m(F(x))}(\bar{F}(x))^m+\frac{f'(x)}{f^2(x)}\Big]. \end{aligned} \tag{13}$$

By definition (10), $g_m(F(x))$ is non-decreasing and in the case $m \neq -1$ its derivative is $f(x)(1-F(x))^m$. Now, by the assumption, $\acute{f}/f^2$ is non-increasing, so the quantity inside the brackets on the right-hand side of Eq. (13) is non-increasing in x and the quantity outside the brackets is positive. Thus $\acute{f}_{X(r,n,m,k)}(x)$ changes sign at most once as x moves from $-\infty$ to ∞ and any change of sign must be from positive to negative. Therefore, $f_{X(r,n,m,k)}(x)$ is unimodal.

(ii) The proof follows similar to that of Theorem 2.2 with $k \in [1, \infty)$. □

It is worth mentioning that only for the case $m = -1$, Theorem 3.1 requires additional condition of f being non-decreasing.

In the next theorem, we present a weaker condition on F than that of Theorem 3.1 (i) for the smallest GOS.

Theorem 3.2. *Suppose that the cdf F with pdf f is unimodal with mode at M. Let $m \neq -1$, f be continuous at M, and $1/f$ be convex in $x < M$. Then, the smallest GOS is unimodal.*

Proof. The pdf of the smallest GOS from (11) is given by

$$f_{X(1,n,m,k)}(x) = \gamma_1(\bar{F}(x))^{\gamma_1-1}f(x), \quad x \in R. \tag{14}$$

Using the same argument as in the Theorem 2.4, we have the result. □

It is worth mentioning that Alam (1972) had shown the validity of above results for ordinary order statistics which is obviously obtained from our Theorems 3.1 (i) and 3.2 when $m = 0$ and $k = 1$.

Finally, we give an analogue of the Theorem 3.1 for the DGOSs.

Theorem 3.3. *Let F be the cdf with pdf f. Then, its DGOS is unimodal if one of the following conditions are satisfied:*
(i) $m \neq -1, 0$, $\frac{1}{f}$ is convex and f is non-decreasing on its support,
(ii) $m = -1, 0$, $\frac{1}{f}$ is convex.

Proof. (i) From (12) we have

$$\begin{aligned} f'_{X_d(r,n,m,k)}(x) &= \frac{c_{r-1}}{(r-1)!}\Big[(\gamma_r-1)f^2(x)(F(x))^{\gamma_r-2}(g_m(\bar{F}(x)))^{r-1} \\ &\quad -(r-1)f^2(x)(F(x))^m(F(x))^{\gamma_r-1}(g_m(\bar{F}(x)))^{r-2} \\ &\quad +f'(x)(F(x))^{\gamma_r-1}(g_m(\bar{F}(x)))^{r-1}\Big] \\ &= \frac{c_{r-1}}{(r-1)!}f^2(x)(F(x))^m(g_m(\bar{F}(x)))^{r-1}(F(x))^{\gamma_r-1} \\ &\quad \times\Big[\frac{(\gamma_r-1)}{(F(x))^{m+1}} - \frac{(r-1)}{g_m(\bar{F}(x))} + \frac{f'(x)}{f^2(x)}\frac{1}{(F(x))^m}\Big]. \end{aligned} \tag{15}$$

Since, in this case, $g_m(\bar{F}(x))$ is non-increasing, the result follows by assumption similarly using the previous arguments.

(ii) For $m = -1$, apply Theorem 2.3 with any rael $k \geq 1$ and for $m = 0$ the result follows by (15). □

Remark 3.1. Similar to Theorem 3.2, for the cases $m = -1, 0$, we have the result for the biggest DGOS by considering the convexity of $1/f$ in $x > M$.

Remark 3.2. Note that, Theorems 3.1 and 3.3 are alike for the case $m = 0$. Thus, for $k = 1$, reversed ordered order statistics have also unimodal distributions by convexity of $1/f$. Also, this holds for order statistics with non-integral sample size by selecting $k = \alpha - n + 1$, where $\alpha \in R$ is the sample size (see Kamps (1995a), p. 51).

Example 3.1. Consider Normal, Cauchy, particular Weibull and Gamma distributions. Then, sequential order statistics with parameters $\alpha_1, \alpha_2, ..., \alpha_n$, $(\alpha_i > 0)$,

$m_i = (n-i+1)\alpha_i - (n-i)\alpha_{i+1} - 1$ and $k = \alpha_n$ for the special case when $(n-i+1)\alpha_i - (n-i)\alpha_{i+1} = c$; and Pfeifer's record model with parameters $\beta_1, \beta_2, ..., \beta_n$, $(\beta_i > 0)$, $m_i = \beta_i - \beta_{i+1} - 1$, and $k = \beta_n$, $(i = 1, 2, ..., n-1)$, (see Kamps (1995a), p. 52-53) for the special case when $\beta_i - \beta_{i+1} = c$, ($c \neq 0$ is a constant), have unimodal distributions in the upper case by Theorem 3.1 (i). Note that, if $c = 0$ and $k \in N$, these models would convert to the k-record model.

References

1. K. Alam, Unimodality of the distribution of an order statistic, *The Annals of Mathematical Statistics*, **43**, 2041-2044, (1972).
2. F. A. Aliev, A comment on 'Unimodality of the distribution of record statistics', *Statistics and Probability Letters*, **64**, 39-40, (2003).
3. B. C. Arnold, N. Balkrishnan and H. N. Nagaraja, *Records*, (Wiley, New York, 1998).
4. P. Basak and I. Basak, Unimodality of the distribution of record statistics, *Statistics and Probability Letters*, **56**, 395-398, (2002).
5. M. Burkschat, E. Cramer and U. Kamps, Dual generalized order statistics, *Metron*, **61**, 13-26, (2003).
6. H. Chen, H. Xie and T. Hu, Log-concavity of generalized order statistics, *Statistics and Probability Letters*, **79**, 396-399, (2009).
7. S. Dharmadhikari and K. Joag-Dev, *Unimodality, Convexity, and Applications*, (Academic Press, Boston, 1988).
8. J. S. Huang and M. Ghosh, A note on strong unimodality of order statistics, (*Journal of the American Statistical Association*, **77**, 929-930, (1982).
9. N. L. Johnson, S. Kotz and N. Balakrishnan, *Continous Univariate Distributions*, Vol. 1. 2nd edn. (Wiley, New York, 1994).
10. N. L. Johnson, S. Kotz and N. Balakrishnan, *Continous Univariate Distributions*, Vol. 2. 2nd edn. (Wiley, New York, 1995).
11. U. Kamps, *A Concept of Generalized Order Statistics*, (Teubner, Stuttgart, 1995a).
12. U. Kamps, A concept of generalized order statistics, *Journal of Statistical Planning and Inference*, **48**, 1-23, (1995b).
13. V. Nevzorov, *Records: Mathematical Theory*, (American Mathematical Society, Providence, RI, 2001).

Nonlinear Model For Time Series Using Mixture Of Autoregressive Time Series With Conditional Heteroscedasticity

S. H. Alizadeh and S. Rezakhah

Department of Statistics, Amirkabir university of Technology

In this paper we introduce a new model for modeling nonlinear time series with heteroscedasticity, and propose a new algorithm for parameter estimation based on EM and fisher scoring iterative algorithms. The MoE-AR-ARCH model is a flexible model for modeling nonlinearities in time series, including burstiness,cycles and flat stretches, since it partitions the covariate space softly into J partitions and fits an autoregressive model with conditional heteroscedasticity in each partition.

Keywords: Nonlinear time series, Mixture Models, Conditional Heteroscedasticity.

1. Introduction

The most frequently used approaches for modeling time series assume that the data under study are generated from a linear Gaussian stochastic process [2]. Linear time series models (such as ARMA, ARMAX and ARIMA) have been well developed and widely used because of their tractability and ease of interpretation. the other reasons for this popularity is that linear Gaussian models provide a number of appealing properties such as frequency domain analysis, asymptotic results, statistical inference and many others that nonlinear models still fail to produce consistently. However, they may be overly simplistic and fail to capture many essential features of the underlying process. It is well known that real-life systems are usually nonlinear, and certain features such as burstiness,limit-cycles, asymmetry, amplitude-dependent frequency responses, jump phenomena, and chaos cannot be correctly captured by linear statistical models [20].

Over recent years, several nonlinear time series models have been proposed both in statistical approach [9] [10] and in machine learning theory [14] [18] [25]. Several nonlinear models have been developed following Tong's seminal work on threshold autoregressive (TAR) models. The central idea of the TAR model is to change the parameters of a linear autoregressive model according to the value of an observable variable, called threshold variable [22]. If this variable is a lagged value of the time series, the model is called self-exciting threshold autoregressive

(SETAR) model. let $\mathbf{x}_t = (y_{t-1}, ..., y_{t-p})^T$. The general Tong's TAR model is expressed as

$$y_t = \sum_{j=1}^{J} (\beta_j + \mathbf{b}_j^T \mathbf{x}_t) I(r_{j-1} \leq y_{t-d} < r_j) + \varepsilon_t$$

Where ε_t are iid zero-mean random variables, $1 < d < p$ represents the delay, r_j represent threshold at which the level and the vector of autoregressive parameters change into new values from those given by β_j and $\mathbf{b}_j$ ($r_0 = -\infty < r_1 \leq ... < r_J = +\infty$), $I(A)$ is the indicator variable ($I(A) = 1$ if A occurs and $I(A) = 0$ otherwise).

Because of difficulties in determining the thresholds empirically, the TAR model is usually is limited to a single threshold r_1 (i.e $J = 2$). the threshold dose not involve interactions among the lagged predictor variables. moreover the regression function of y_t on $\mathbf{x}_t$ has jump discontinuities [16]. A natural generalization of the SETAR model is the STAR model, proposed by Chan and Tong [5] and expressed as

$$y_t = \phi_0^{(1)} + \sum_{i=1}^{P} \phi_0^{(1)} y_{t-i} + (\phi_0^{(2)} + \sum_{i=1}^{P} \phi_0^{(2)} y_{t-i}) F(\gamma y_{t-d} - r) + \varepsilon_t \qquad (1)$$

where $F(.)$, called transition function, is a continuous, monotonically increasing function. The parameter γ is responsible by the smoothness of the function $F(.)$. When $\gamma \to \infty$, (1) becomes a SETAR model with two regimes. The scalar parameter is known as the location parameter. Although this model circumvents the problem of jump discontinuities, but the number of autoregressive components in the model is restricted to 2 and the other problems of Tong's TAR model is still remained [20]. Medeiros and A. Veiga extended STAR model with the aid of multi layer perceptron (MLP) neural network. They proposed neuro-coefficient smooth transition autoregressive (NCSTAR) model for forecasting nonlinear time series. The restriction on the number of partitions was remedied in Their model, the neurons in the hidden layer of MLP enables the model to partition the covariate space along arbitrary hyperplanes [20]. Lai and Wong extended the tong's TAR model with the aid of stochastic neural networks [16]. Stochastic neural network was used as a method for soft splitting the covariate space.

Another approach for modeling nonlinear time series consist of modeling conditional distributions of time series. This idea originated from seminal work of Raftery for modeling high order markov chains [21]. Raftery et. al. extended the mixture transition distribution (MTD) model for continuous state space processes and introduced GMTD model [17]. GMTD was able to model flat stretches, bursts and outliers in time series. Wong and Li further extended GMTD model to Mixture of Autoregressive (MAR) model [23]. Berchtold used MTD model with time

varying variance components to model heteroscedastic time series [3]. Amendola and Nigli derived the predictor distribution and for cast accuracy of threshold models, they showed that the conditional density of time series for multi step ahead prediction of threshold models follows a finite mixture model [1]. A special case of finite mixture model in which both the mixture coefficients and the mixture components are generalized linear model(GLIM's) was proposed by Jordan and Jacobs, they called this model Mixture of Experts (MoE) neural network and derived an efficient algorithm for parameter estimation [12] [13]. In fact they used the idea of mixture models in conjunction with soft splitting property of neural networks for decreasing the variance of estimator. Carvalho and Tanner used Mixtures of Experts of Autoregressive Time Series for modeling nonlinear time series, they derived the necessary conditions for asymptotic normality of MLE estimator of their model.

In this paper we introduce a new model for modeling nonlinear time series with heteroscedasticity, and propose a new algorithm for parameter estimation based on EM and fisher scoring iterative algorithms. The MoE-AR-ARCH model is a flexible model for modeling nonlinearity in time series, including burstiness,cycles and flat stretches, since it partitions the covariate space softly into J partitions and fits an autoregressive model with conditional heteroscedasticity in each partition.

2. Mixtures-of-Experts of Autoregressive Time Series with Conditional Heteroscedasticity

The J-component Mixtures-of-Experts of Autoregressive Time with Conditional Heteroscedasticity (MoE-AR-ARCH) model under consideration is defined by

$$\begin{aligned} F(y_t|\mathfrak{F}_{t-1}) &= F(y_t|\mathbf{y}_{t-1},\theta), \\ &= \sum_{j=1}^{J} g_j(\mathbf{y}_{t-1},\gamma)F_j(y_t|\mathbf{y}_{t-1},\theta_j) \\ &= \sum_{j=1}^{J} g_j(\mathbf{y}_{t-1},\gamma)\Phi(\frac{y_t-\phi_j'\mathbf{y}_{t-1}}{\sigma_{j,t}}), \end{aligned} \tag{2}$$

and the weighting function $g_j(.,\gamma)$ is defined by

$$g_j(\mathbf{y}_{t-1},\gamma) = \frac{e^{\beta_j'\mathbf{y}_{t-1}}}{\sum_{i=1}^{J} e^{\beta_i'\mathbf{y}_{t-1}}} \tag{3}$$

and the time varying variance (Heteroscedasticity) of each component is defined by

$$\begin{aligned} \sigma_{j,t}^2 &= \alpha_{j_0} + \alpha_{j1}\epsilon_{j,t-1}^2 + ... + \alpha_{jq_j}\epsilon_{j,t-q_j}^2 \\ \epsilon_{j,t} &= y_t - \phi_{j0} - \phi_{j1}y_{t-1} - ... - \phi_{jp_j}y_{t-p_j} \end{aligned} \tag{4}$$

where $\Phi(.)$ is the distribution of standard normal random variable, $\mathbf{y}_{t-1} = (1, y_{t-1}, ..., y_{t-p})'_{p\times 1}$, $\phi_j = (\phi_{j0}, \phi_{j1}, ..., \phi_{jp_j}, 0, ..., 0)'_{p\times 1}$, $\alpha_j = (\alpha_{j_0}, \alpha_{j1}, ..., \alpha_{jq_j})'_{p\times 1}$, $\beta_j = (\beta_{j0}, \beta_{j1}, ..., \beta_{jp})'_{p\times 1}$ and $p = max\{p_j, j = 1, ...J\}$. From equation (3) it is clear that weighting functions $g_j(\mathbf{y}_{t-1}) \in (0, 1)$ and $\sum_{j=1}^{J} g_j(\mathbf{y}_t) = 1$, thus the conditional distribution in model (2) is convex combination of J distinct distributions, As required by finite mixture models. The weighting function is a function of lagged time series variables, this model reduces to finite mixture model of autoregressive time series with conditional heteroscedasticity if it is constant (i.e. the MAR-ARCH model proposed by Wong and Li in 2001 [**?**]). Note that the order of autoregressive components are not necessarily the same, we add the required number of zeros at the end of original ϕ_j to rise it's size up to p. To avoid the possibility of zero or negative conditional variance, the following conditions for α_{jk} must be imposed

$$\begin{aligned} &\alpha_{j0} > 0 \\ &\alpha_{ji} \geq 0 \quad j = 1, ..., J, i = 1, ..., q_j \end{aligned}$$

The core idea of MoE-AR-ARCH model for modeling nonlinear heteroscedastic time series is based on the divide-and-conquer approach for solving nonlinear complex problems [13] and Engle's idea for modeling heteroscedasticity in time series. At any time point we have a certain number of heteroscedastic autoregressive models, we assume the existence of latent multinomial random variable which determines the probability of associated model for generating the observations of time series. We adopt the multinomial logit models for modeling the distribution of latent variable. In other words the weighting functions (3) determine the probability that the next observation of time series comes from jth component of mixture model. In fact the weighting functions softly split the covarite space into J partitions. By soft split, we meaning that data points may lie simultaneously in multiple regions. On the other hand for real processes one might expect better forecast intervals if additional information from the past were allowed to affect the forecast variance, for this end we made use of Engle's ARCH model [8] in MoE-AR-ARCH model.

In MoE-AR-ARCH model, the conditional expectation of y_t given the past observations $y_{t-i}, i > 1$ is

$$E[y_t|\mathfrak{F}_{t-1}] = E[y_t|\mathbf{y}_{t-1}] = \sum_{j=1}^{J} g_j(\mathbf{y}_{t-1}, \gamma)\phi'_j\mathbf{y}_{t-1} \tag{5}$$

and its conditional variance is

$$\begin{aligned} Var(y_t|\mathfrak{F}_{t-1}) &= Var(y_t|\mathbf{y}_{t-1}) \\ &= \sum_{j=1}^{J} g_j(\mathbf{y}_{t-1},\gamma)\sigma_{j,t}^2 + \sum_{j=1}^{J} g_j(\mathbf{y}_{t-1},\gamma)(\phi_j'\mathbf{y}_{t-1})^2 \\ &\quad -(\sum_{j=1}^{J} g_j(\mathbf{y}_{t-1},\gamma)\phi_j'\mathbf{y}_{t-1})^2 \end{aligned} \tag{6}$$

3. Estimation

Estimation of the parameters of MoE-AR-ARCH model,$\Theta = \{\theta_j = \{\phi_j, \alpha_j\}, \beta_j, j = 1...J\}$ is similar to finite mixture models but with more complications. Most of this complexity is due to the components heteroscedasticity. The EM algorithm ,the most readily available procedure in estimating mixture type models, is used.

The EM algorithm is a general iterative technique for maximum likelihood estimation. Each iteration is composed of two steps: A estimation (**E**) step and a maximization (**M**) step. An application of EM generally begins with the observation that the optimization of the likelihood function $l(\Theta, \mathbf{Y})$,$\mathbf{Y} = (y_1, ..., y_n)$ would be simplified if only a set of additional variables, called "missing" or "hidden" variables, were known. We refer to observable data $\mathbf{Y}$ as the incomplete data and point a "complete data" set $\mathbf{Y}^c$ that includes missing variable $\mathbf{Z}$ and observed data $\mathbf{Y}$. First calculate the expected value of the complete-data likelihood($l^c(\Theta, \mathbf{Y}^c)$), given the observed data and the current model, This is the E step:

$$Q(\Theta, \Theta^{(p)}) = E[l^c(\Theta, \mathbf{Y}^c)|\mathbf{Y}] \tag{7}$$

where $\Theta^{(p)}$ is the value of parameter at pth iteration and the expectation is taken with respect to it. this step yields a deterministic function Q. The M step maximizes this function with respect to Θ to find the new parameter estimates $\Theta^{(p+1)}$

$$\Theta^{(p+1)} = \arg\max_{\Theta} Q(\Theta, \Theta^{(p)}) \tag{8}$$

Dempster proved that an increase in Q implies an increase in the incomplete likelihood [11].

$$l(\Theta^{(p+1)}, \mathbf{Y}) \geq l(\Theta^{(p)}, \mathbf{Y}^c) \tag{9}$$

Suppose that the Observations $\mathbf{Y} = (y_1, ..., y_n)$ is generated from the model (2), let $\mathbf{Z} = (Z_1, Z_2, ..., Z_n)$ be the missing data where $Z_t = (z_{1,t}, ..., z_{J,t})$ is a J-dimensional unobservable random vector with it's jth component equal to 1 if

y_t comes from the jth component of the conditional distribution function and 0 otherwise.

$$f(y_t, z_t|\mathbf{y}_{t-1}, \Theta) = \prod_{j=1}^{J} (g_j(\mathbf{y}_{t-1}) f_j(y_t|\mathbf{y}_{t-1}, \theta_j))^{z_{j,t}} \tag{10}$$

The partial likelihood of the complete data is written as:

$$\begin{aligned} l^c(\theta, \mathbf{Y}^c) &= \sum_{t=1}^{n} \sum_{j=1}^{J} z_{j,t} ln(g_j(\mathbf{y}_{t-1}) f_j(y_t|\mathbf{y}_{t-1}, \theta_j)) \\ &= \sum_{t=1}^{n} \sum_{j=1}^{J} z_{j,t} \{ln(g_j(\mathbf{y}_{t-1})) + ln(f_j(y_t|\mathbf{y}_{t-1}, \theta_j))\} \end{aligned} \tag{11}$$

$$E[z_{j,t}|\mathbf{Y}] = P(z_{j,t} = 1|y_t, \mathbf{y}_{t-1}, \Theta^{(p)}) \tag{12}$$

Using the Bayes's rule, this prior probability can be computed based on the posterior probabilities. The posterior probability is defined once both the input and the target output are known.

$$P(z_{j,t} = 1|y_t, \mathbf{y}_{t-1}, \Theta^{(p)}) = h_{j,t} = \frac{g_j(\mathbf{y}_{t-1}) f_j(y_t|\mathbf{y}_{t-1}, \theta_j)}{\sum_{i=1}^{J} g_i(\mathbf{y}_{t-1}) f_i(y_t|\mathbf{y}_{t-1}, \theta_j)} \tag{13}$$

Using previous equation and the complete data log likelihood (11), the conditional expectation of complete data log likelihood (7) can be written az:

$$Q(\Theta, \Theta^{(p)}) = \sum_{t=1}^{n} \sum_{j=1}^{J} h_{j,t} \{ln(g_j(\mathbf{y}_{t-1})) + ln(f_j(y_t|\mathbf{y}_{t-1}, \theta_j))\} \tag{14}$$

The **M** step requires maximizing $Q(\Theta, \Theta^{(p)})$ with respect to parameters set Θ:

$$\theta_j^{(p+1)} = \arg\max_{\theta_j} \sum_t h_{j,t} ln(f_j(y_t|\mathbf{y}_{t-1}, \theta_j)) \tag{15}$$

$$\beta_j^{(p+1)} = \arg\max_{\beta_j} \sum_t \sum_j h_{j,t} ln(g_j(\mathbf{y}_{t-1})) \tag{16}$$

Equation (16) involves maximizing the cross-entropy between posterior probabilities $h_{j,t}$ and the prior probabilities, $g_j(\mathbf{Y}_{t-1})$. this cross-entropy is the log likelihood associated with a multinomial logit probability model, in which the $h_{j,t}$ act as the output observations. Thus this is a maximum likelihood problem and can be solved using IRLS [13]. nonlinear optimization of equation (15) is done with

an iterative algorithm. The iterations are calculated using scoring algorithm. Each step for parameter vector θ produces estimates θ_j^{i+1} based on θ_j^i according to

$$\theta_j^{i+1} = \theta_j^i + [\hat{\Im}_{\theta_j\theta_j}]^{-1}\frac{1}{T}\sum_t \frac{\partial l_t^i}{\partial \theta_j} \tag{17}$$

where $\hat{\Im}$ and $\partial l_t^i/\partial\theta_j$ are evaluated at θ_j^i.

Because of the components heteroscedasticity, direct calculation of (17) is so complected, since

$$\frac{\partial l_t^i}{\partial \phi_j} = \frac{1}{T}\{\sum_{t=1}^{T}\frac{h_{j,t}}{2\sigma_{j,t}^2}\frac{\partial \sigma_{j,t}^2}{\partial \phi_j}(\frac{\epsilon_{j,t}^2}{\sigma_{j,t}^2}-1) - \sum_{t=1}^{T}\frac{h_{j,t}\epsilon_{j,t}}{2\sigma_{j,t}^2}\frac{\partial \epsilon_{j,t}}{\partial \phi_j}\} \tag{18}$$

where

$$\frac{\partial \sigma_{j,t}^2}{\partial \phi_{j,k}} = -\sum_{i=1}^{k} 2\alpha_{j,i}\epsilon_{j,t-i}y_{t-k} \tag{19}$$

Lemma 3.1. *Because of the block diagonality of the information matrix, the estimation of ϕ_j and α_j can be considered separately without loss of asymptotic efficiency. Furthermore either can be estimated with full efficiency based only on a consistent estimate of the other where $h_{j,t}$ is defined in equation (13).*

Proof. see Engle's Article "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation", [8]. □

Using the previous lemma, The iterative scoring algorithm for estimation of parameters of MoE-AR-ARCH model in the **M** step of EM algorithm can be summarized in the following steps:

1. As it is can be seen from equation (15), optimizing of equation (14) involves J distinct optimization problem for determining the parameters of each component separately.
2. Initially estimate ϕ_j by ordinary least squares, and obtain residuals.
3. From the residuals of previous step, an efficient estimate of α_j can be constructed.
4. Based on the estimates, $\hat{\alpha}_j$, efficient estimates of ϕ_j are found.

Note that in the proposed algorithm, $\frac{\partial l_t^i}{\partial \phi_j}$ in equation (18) only contains the second term.

4. conclusion

In this paper we introduced a new model for modeling nonlinear time series with heteroscedasticity, and proposed a new algorithm for parameter estimation based on EM and fisher scoring iterative algorithms. The MoE-AR-ARCH model is a flexible model for modeling nonlinearity in time series, including burstiness,cycles and flat stretches, since it partitions the covariate space softly into J partitions and fits an autoregressive model with conditional heteroscedasticity in each partition.

References

1. A. Amendola, M. Niglio, "Pridiction distribution and forecast accuracy of threshold models", *Statistical Methods & Applications*, Vol. 13, pp. 1-12,2004.
2. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis, Forecasting and Control*, **3rd ed.**, Englewood Cliffs, NJ: Prentice-Hall, 1994.
3. A. Berchtold, "Mixture transition distribution modeling of heteroscedastic time series", *Computational Statistics and Data Analysis*, No. 41, pp. 399-411,2003.
4. , A.X. Carvalho, M.A. Tanner, "Mixtures-of-Experts of Autoregressive Time Series:Asymptotic Normality and Model", Specification", *IEEE Transaction on Neural Networks*,Vol 16, No 1,2005.
5. K. S. Chan and H. Tong, On estimating thresholds in autoregressive models, *J. Time Series Anal.*, vol. 7, pp. 179190,1986.
6. D.R. Cox, and Hinkley, *Theoretical Statistics.*, London: Chapman and Hall,1974.
7. A. P. Dempster, N. M. Larid, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm"(with discussion), *Journal of the Royal statistical society*, Ser B, 39, 1-18,1977.
8. R. F. Engle, "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation,", *Econometrica*, 50,pp. 987-1007,1982.
9. J. Fan and Q. Yao, *Nonlinear Time Series:Nonparametric and parametric methods*, New York, Springer Series in Statistics,2003.
10. J. Gao, *Nonlinear Time Series Semiparametric and Nonparametric Methods*,6000 Broken Sound Parkway NW, Suite 300,Chapman & Hall/CRC,2007.
11. G.H. Givens, J.A. Hoeting, ,*Computational statistics*, Wiely, United state of America,2004.
12. R.A. Jacobs, M.I. Jordan, "Adaptive Mixture of local experts.", *Neural Computation*, vol. 3, pp. 79-87,1991.
13. M.I. Jordan , R. A. Jacobs, "Hierarchical Mixtures of Experts and the EM Algorithm". *Tech. Rep* ,1993.
14. G. Kirchgassner and J. Wolters, *Introduction to Modern Time Series Analysis*, New York ,Springer,2007.
15. B. Keden,K. Fokianos, *Regression Models for Time Series Analysis*, WILEY, United States of America,2002.
16. T.L. Lai and S.P. Wong, " Stochastic Neural Networks with Applications to Nonlinear Time Series". *J. Am. Statist. Assoc.*, vol. 96, No 455,2001.
17. N.D. Le, R.D. Martin, A.E. Raftery, "Modeling flat stertches,bursts and outliers in time series using mixture transition distribution models",*J. Amer. Statist. Assoc. 91*, pp. 1504-15,1996.

18. P.D. McNelis, *Neural Networks in Finance Gaining Predictive Edge in the Market*, New York,Elsevier Academic Press,2005.
19. M. McCullagh and J. Nelder, *Generalized Linear Models* (Monographs on Statistics and Applied Probability, 37). London, U.K., Chapman & Hall, 1998.
20. M. C. Medeiros and A. Veiga, "A Hybrid Linear-Neural Model for Time Series Forecasting". *IEEE TRANSACTIONS ON NEURAL NETWORKS*, VOL. 11, NO. 6, pp. 1402-12,2000.
21. A.E. Raftery, "A model for high order Markov chains,", *J. Roy. Statistics Co.* B. 47, pp. 528-539,1985.
22. H. Tong, K.S. Lim, "Threshold Autoregression, Limit Cycle and Cyclical Data"(with discussion), *Journal of the Royal Statistical Society*, Ser. B, 42, pp. 245292,1980.
23. C.S. Wong, W.K. Li, "On a Mixture Autoregressive Model", *J. Roy. Statist. Soc.*, B 62, pp.92-115,2000.
24. M. Villani , R. Kohnc, P. Giordani, "Regression density estimation using smooth adaptive Gaussian mixtures",*Journal of Econometrics*,2009.
25. G. Zhang, B. E. Patuwo, and M.Y. Hu, "Forecasting with artificial neural networks: The state of the art", *Int. J. Forecasting*, vol. 14, pp. 3562, 1998.

The asymptotic efficiencies of the MLE based on bivariate record values from bivariate normal distribution

M. Aminia and J. Ahmadi

Department of Statistics, Ferdowsi University of Mashhad

One of the most important bivariate distributions in statistical inference is the bivariate normal distribution. In this paper we consider ML estimation of parameters of bivariate normal distribution using bivariate record values. The asymptotic properties especially asymptotic efficiencies of the ML estimators are studied and calculated by deriving the Fisher information matrix about the five dimension parameter vector contained in the vector of first n bivariate record values. The relative asymptotic efficiencies are concerned with respect to ML estimators based on a bivariate sample of size n and the corresponding ML estimators based on bivariate record values and inter-record times.

Keywords: Additivity; Bivariate distribution; inverse sampling.

1. Introduction and preliminaries

Let $\{(X_i, Y_i), i \geq 1\}$ be a sequence of i.i.d. pair-wise random variables with the absolutely continuous cumulative distribution function (cdf) $F_{(X,Y)}(x, y; \theta)$ and the corresponding pdf $f_{(X,Y)}(x, y; \theta)$. Also $f_X(x; \theta)$ and $F_X(x; \theta)$ denote the marginal pdf and cdf of Xs, respectively and $\bar{F}_X(x; \theta) = 1 - F_X(x; \theta)$. The sequence of upper records and their concomitants is defined as $(R_n, R_{[n]}) = (X_{T_n}, Y_{T_n}), \quad n \geq 1$, where $T_1 = 1$ with probability one and for $n \geq 2$, $T_n = \min\{j : j > T_{n-1},\ X_j > X_{T_{n-1}}\}$. Suppose $\Delta_i = T_{i+1} - T_i - 1,\ i = 1, 2, \ldots, n-1$, $\Delta_n = 0$ are the number of trials needed to obtain new records, which are called inter-record times. Let us denote $\mathbf{R}_n = (R_1, \ldots, R_n),\ \mathbf{\Delta}_n = (\Delta_1, \ldots, \Delta_n), \mathbf{C}_n = (R_{[1]}, \ldots, R_{[n]})$.

The sequence of records and concomitants are called bivariate records. The joint pdf of the first n bivariate upper records is (See Arnold et al., 1998)

$$f_{(\mathbf{R}_n, \mathbf{C}_n)}(\mathbf{r}_n, \mathbf{r}_{[n]}) = \prod_{i=1}^{n} f_{(X,Y)}(r_i, s_i; \theta) / \prod_{i=1}^{n-1} [\bar{F}_X(r_i; \theta)]. \tag{1}$$

Using (1) the joint pdf of records, inter-record times and their concomitants is

given by

$$f_{(\mathbf{R}_n,\mathbf{\Delta}_n,\mathbf{C}_n)}(\mathbf{r},\delta,\mathbf{s};\theta)=\prod_{i=1}^{n} f_{X,Y}(r_i,s_i;\theta)\{F_X(r_i;\theta)\}^{\delta_i}. \tag{2}$$

So, the conditional probability mass function of $\mathbf{\Delta}_n$ given $(\mathbf{R}_n,\mathbf{C}_n)$ is given by

$$f_{(\mathbf{\Delta}_n|\mathbf{R}_n,\mathbf{C}_n)}(\delta|\mathbf{r},\mathbf{s};\theta)=\prod_{i=1}^{n-1}[F_X(r_i;\theta)]^{\delta_i}\bar{F}_X(r_i;\theta). \tag{3}$$

We consider upper bivariate records to derive the results of this paper. Since for standard normal distribution, we have R'_n (nth lower record) and $-R_n$ are identically distributed, similar results also hold for the case of lower records. Hereafter, we will call upper records, simply, records. We denote some moments of bivariate records from standard bivariate normal distribution as follows $\alpha_{i[i]}=\mathrm{E}(R_i^{0,1}R_{[i]}^{0,1})$, $\alpha_{[i]}^{(2)}=\mathrm{E}(R_{[i]}^{0,1})^2$ and $\alpha_i^{(2)}=\mathrm{E}(R_i^2)$, where $(R_i^{0,1},R_{[i]}^{0,1})$ is the ith bivariate record from standard normal distribution. Suppose the distribution of (X_i,Y_i) is standard bivariate normal with correlation ρ for i=1,2,..., then $Y_i=\rho X_i+\epsilon_i$, where the X_i and the ϵ_i are mutually independent and ϵ_i are normal distributed with zero mean and variance equal to $1-\rho^2$. So by considering X-record sequence we have for $n\geq 0$, $R_{[i]}^{0,1}=\rho R_i^{0,1}+\epsilon_{[i]}$, where $\epsilon_{[i]}$ denotes the particular ϵ_i associated with $R_i^{0,1}$. Since X_i and ϵ_i are independent so the sequence $R_i^{0,1}$ is independent of $\epsilon_{[i]}$, the later being mutually independent, each with the same distribution as ϵ_i. So we can conclude that $\alpha_{i[i]}=\rho\alpha_i^{(2)}$ and $\alpha_{[i]}^{(2)}=\rho^2\alpha_i^{(2)}+1-\rho^2$.

2. ML estimation

Suppose that m independent sequences each with k bivariate records are observed and the sampling distribution is bivariate normal distribution with parameter $\theta=(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)$. Let $(R_{i(j)},R_{[i](j)})$ denote the ith bivariate record in jth sequence. Our aim is to obtain the maximum likelihood estimator of θ. The likelihood equations for bivariate record values and record times are as follows:

$$\sum_{j=1}^{m}\sum_{i=1}^{k}\left(\frac{R_{i(j)}-\mu_1}{\sigma_1}\right)-\frac{\rho}{\sigma_1}\sum_{j=1}^{m}\sum_{i=1}^{k}\left(\frac{R_{[i](j)}-\mu_2}{\sigma_2}\right)$$
$$-\frac{1-\rho^2}{\sigma_1}\sum_{j=1}^{m}\sum_{i=1}^{k}\Delta_i r_0\left(\frac{R_{i(j)}-\mu_1}{\sigma_1}\right)=0, \tag{4}$$

$$\sum_{j=1}^{m}\sum_{i=1}^{k}R_{[i](j)}-mk\mu_2-\frac{\rho\sigma_2}{\sigma_1}\left(\sum_{j=1}^{m}\sum_{i=1}^{k}R_{i(j)}-mk\mu_1\right)=0,$$

$$\rho\,(mk - \frac{\sum_{j=1}^{m}\sum_{i=1}^{k}(R_{i(j)}-\mu_1)^2}{\sigma_1^2} - \frac{\sum_{j=1}^{m}\sum_{i=1}^{k}(R_{[i](j)}-\mu_2)^2}{\sigma_2^2}\Big)$$
$$+ C^*(\theta)(\rho^2+1) - mk\rho^3 = 0,$$

where $C^*(\theta) = \sum_{j=1}^{m}\sum_{i=1}^{k}\left(R_{i(j)}-\mu_1\right)\left(R_{[i](j)}-\mu_2\right)/(\sigma_1\sigma_2)$,

$$mk(1-\rho^2) - \sum_{j=1}^{m}\sum_{i=1}^{k}\left(\frac{R_{i(j)}-\mu_1}{\sigma_1}\right)^2 + \rho C^*(\theta)$$
$$+ (1-\rho^2)\sum_{j=1}^{m}\sum_{i=1}^{k}\frac{\Delta_i(R_{i(j)}-\mu_1)}{\sigma_1} r_0\left(\frac{R_{i(j)}-\mu_1}{\sigma_1}\right) = 0, \qquad (5)$$

and

$$mk(1-\rho^2) - \sum_{j=1}^{m}\sum_{i=1}^{k}\left(\frac{R_{[i](j)}-\mu_2}{\sigma_2}\right)^2 + \rho C^*(\theta) = 0.$$

For the case of bivariate record values without record times, the above equations remain true except that (4) and (5) are replaced, respectively, with

$$\sum_{j=1}^{m}\sum_{i=1}^{k}\left(\frac{R_{i(j)}-\mu_1}{\sigma_1}\right)$$
$$- \frac{\rho}{\sigma_1}\sum_{j=1}^{m}\sum_{i=1}^{k}\left(\frac{R_{[i](j)}-\mu_2}{\sigma_2}\right)$$
$$- \frac{1-\rho^2}{\sigma_1}\sum_{j=1}^{m}\sum_{i=1}^{k-1} h_0\left(\frac{R_{i(j)}-\mu_1}{\sigma_1}\right) = 0,$$

and

$$mk(1-\rho^2) - \sum_{j=1}^{m}\sum_{i=1}^{k}\left(\frac{R_{i(j)}-\mu_1}{\sigma_1}\right)^2 + \rho C^*(\theta)$$
$$+ (1-\rho^2)\sum_{j=1}^{m}\sum_{i=1}^{k-1}\frac{(R_{i(j)}-\mu_1)}{\sigma_1} h_0\left(\frac{R_{i(j)}-\mu_1}{\sigma_1}\right) = 0.$$

The roots of likelihood equations have no closed form and these equations have to be solved numerically.

3. Asymptotic properties

In this section, we study the consistency and asymptotic normality of the maximum likelihood estimator of θ, obtained in the previous section.

Theorem 3.1. *Let $\hat{\theta}_m$ be the MLE of θ from the bivariate normal distribution based on m independent sequences of the first k bivariate record values, then $\hat{\theta}_m$ is a consistent estimator for θ as $m \to \infty$.*

Proof. Let θ_0 be the value of θ. Since $\hat{\theta}_m$ is the ML estimator of θ, we have

$$\log\ f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\hat{\theta}_m) \geq \log\ f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta), \quad \text{for all } \theta \in \Theta. \tag{6}$$

On the other hand by Jensen's inequality we have

$$\mathrm{E}_{\theta_0}\left(\log \frac{f_{\mathbf{R},\mathbf{C}}(\mathbf{r},\mathbf{s};\theta)}{f_{\mathbf{R},\mathbf{C}}(\mathbf{r},\mathbf{s};\theta_0)}\right) \leq \log \mathrm{E}_{\theta_0}\left(\frac{f_{\mathbf{R},\mathbf{C}}(\mathbf{r},\mathbf{s};\theta)}{f_{\mathbf{R},\mathbf{C}}(\mathbf{r},\mathbf{s};\theta_0)}\right) = 0.$$

So

$$\mathrm{E}_{\theta_0}\left(\log\ f_{\mathbf{R},\mathbf{C}}(\mathbf{r},\mathbf{s};\theta)\right) \leq \mathrm{E}_{\theta_0}\left(\log\ f_{\mathbf{R},\mathbf{C}}(\mathbf{r},\mathbf{s};\theta_0)\right).$$

Since $\frac{1}{m}\log\ f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)$ tends almost everywhere to $\mathrm{E}_{\theta_0}\left(\log\ f_{\mathbf{R},\mathbf{C}}(\mathbf{r},\mathbf{s};\theta)\right)$ as $m \to \infty$, hence with probability one and for large enough m, we have

$$\frac{1}{m}\log\ f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta) \leq \frac{1}{m}\log\ f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta_0), \quad \text{for all } \theta \in \Theta. \tag{7}$$

Setting $\theta = \theta_0$ in (6) and $\theta = \hat{\theta}_m$ in (7), it is deduced that $\hat{\theta}_m$ tends to θ_0 with probability one, as $m \to \infty$. □

Theorem 3.2. *Let $\hat{\theta}_m$ be as in Theorem 3.1, then the asymptotic distribution of $\sqrt{m}(\hat{\theta}_m - \theta)$ is $\mathcal{N}_5(\mathbf{0}\ ,\ I^{-1}_{\mathbf{R},\mathbf{C}}(\theta))$ as $m \to \infty$, where $I^{-1}_{\mathbf{R},\mathbf{C}}(\theta)$ is the inverse of Fisher information matrix in $(\mathbf{R},\mathbf{C})$.*

Proof. The Taylor's expansion of $\frac{\partial}{\partial\theta_r}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\Big|_{\theta=\hat{\theta}}$ for $r = 1,\ldots,k$ around an arbitrarily θ_1 is equal to

$$\begin{aligned}\frac{\partial}{\partial\theta_r}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\Big|_{\theta=\hat{\theta}} &= \frac{\partial}{\partial\theta_r}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\Big|_{\theta=\theta_1} \\ &\quad + \sum_{s=1}^{5}(\hat{\theta}_s - \theta_0)\ \frac{\partial^2}{\partial\theta_r\partial\theta_s}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\Big|_{\theta=\theta^*},\end{aligned}$$

where θ^* is a sequence in Θ which tends in probability to θ_1. The left hand side of the above equality is equal to zero and hence

$$\begin{aligned}\frac{\partial}{\partial\theta}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\Big|_{\theta=\theta_1} &= \sum_{s=1}^{5}(\hat{\theta}_s - \theta_1)\ \frac{-\partial^2}{\partial\theta\partial\theta_s}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\Big|_{\theta=\theta^*} \\ &= m(\hat{\theta} - \theta_1)I_{\mathbf{R},\mathbf{C}}(\theta).\end{aligned}$$

On the other hand, $\frac{\partial}{\partial\theta}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\big|_{\theta=\theta_1}$ has a mean equal to zero and a variance equal to $mI_{\mathbf{R},\mathbf{C}}(\theta)$. So

$$\frac{1}{\sqrt{m}}\,\frac{\partial}{\partial\theta}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\bigg|_{\theta=\theta_1}=\sqrt{m}\left(\frac{1}{m}\,\frac{\partial}{\partial\theta}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\bigg|_{\theta=\theta_1}-0\right)$$

has an asymptotic distribution $\mathcal{N}_5(\mathbf{0}\ ,\ I_{\mathbf{R},\mathbf{C}}(\theta))$ as $m\to\infty$. Hence,

$$\sqrt{m}(\hat{\theta}_m-\theta)=I^{-1}_{\mathbf{R},\mathbf{C}}(\theta)\frac{1}{\sqrt{m}}\,\frac{\partial}{\partial\theta}\log f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)\bigg|_{\theta=\theta_1}$$

has an asymptotic distribution $\mathcal{N}_5(\mathbf{0}\ ,\ I^{-1}_{\mathbf{R},\mathbf{C}}(\theta))$ as $m\to\infty$. □

Remark: Similar asymptotic results are hold for ML estimators based on bivariate records and inter-record times by replacing $f_{\mathcal{R}_m,\mathcal{C}_m}(\varrho_m,\varsigma_m;\theta)$ with $f_{\mathcal{R}_m,\mathcal{C}_m,\nabla_m}(\varrho_m,\varsigma_m,\delta_m;\theta)$ and $I_{\mathbf{R},\mathbf{C}}(\theta)$ with $I_{\mathbf{R},\mathbf{C},\mathbf{\Delta}}(\theta)$ in Theorems 3.1 and 3.2.

4. Fisher information matrix

In order to compute the asymptotic relative efficiencies of the estimator of θ based on m independent sequences each with k bivariate records and times (denoted with RCT), with corresponding estimator based on similar mk bivariate records without times (denoted with RC), and the estimator based on a bivariate random sample of size mk (denoted with IID), we derive the Fisher information matrices of these three data bases. We have $kI_{X,Y}(\theta)=((I_{ij}));\ i,j=1\ldots,5$, such that $I_{ij}=I_{ji};\ j\neq i$ and $I_{11}=\frac{k}{\sigma_1^2(1-\rho^2)},\ I_{12}=0,\ I_{13}=\frac{-k\rho}{\sigma_1\sigma_2(1-\rho^2)},\ I_{14}=0,\ I_{15}=0,\ I_{22}=\frac{k(1-\rho^2/2)}{2\sigma_1^4(1-\rho^2)},\ I_{23}=0,\ I_{24}=\frac{-k\rho^2}{4\sigma_1^2\sigma_2^2(1-\rho^2)},\ I_{25}=\frac{-k\rho}{2\sigma_1^2(1-\rho^2)},\ I_{33}=\frac{k}{\sigma_2^2(1-\rho^2)},\ I_{34}=0,\ I_{35}=0,\ I_{44}=\frac{k(1-\rho^2/2)}{2\sigma_2^4(1-\rho^2)},\ I_{45}=\frac{-k\rho}{2\sigma_2^2(1-\rho^2)},$ and $I_{55}=\frac{k(1+\rho^2)}{(1-\rho^2)^2}$. Let

$$L^*=L^*(\theta;r_i,s_i)=-\frac{1}{2}\log(\sigma_1^2\sigma_2^2(1-\rho))-\frac{1}{2(1-\rho^2)}\left[\left(\frac{r_i-\mu_1}{\sigma_1}\right)^2+\left(\frac{s_i-\mu_2}{\sigma_2}\right)^2-\ 2\rho\left(\frac{r_i-\mu_1}{\sigma_1}\right)\left(\frac{s_i-\mu_2}{\sigma_2}\right)\right].$$

Hence, we obtain $I_{RC}(\theta)=((I'_{ij}));\ i,j=1\ldots,5$, such that $I'_{ij}=I'_{ji};\ j\neq i$ and

$$\begin{aligned}I'_{11}&=-\sum_{i=1}^{k}\mathrm{E}\left(\frac{\partial^2}{\partial\mu_1^2}L^*(\theta;R_i,R_{[i]})\right)-\frac{1}{\sigma_1^2}\sum_{i=1}^{k-1}\mathrm{E}(h'_0(R_i^{0,1}))\\&=\frac{1}{\sigma_1^2}\left[\frac{k}{1-\rho^2}+\sum_{i=1}^{k-1}\mathrm{E}(R_i^{0,1}h_0(R_i^{0,1}))-\sum_{i=1}^{k-1}\mathrm{E}(h_0^2(R_i^{0,1}))\right],\end{aligned}$$

$$I'_{22} = -\sum_{i=1}^{k} \mathrm{E}\left(\frac{\partial^2}{\partial(\sigma_1^2)^2} L^*\right) - \frac{1}{4\sigma_1^4}\mathrm{E}\left[3\sum_{i=1}^{k-1} R_i^{0,1} h_0(R_i^{0,1}) + \sum_{i=1}^{k-1}(R_i^{0,1})^2 h_0'(R_i^{0,1})\right]$$
$$= \frac{1}{\sigma_1^4}\left[\frac{-k}{2} + \frac{1-3/4\rho^2}{(1-\rho^2)}\sum_{i=1}^{k}\alpha_i^{(2)} - \frac{3}{4}\sum_{i=1}^{k-1}\mathrm{E}(R_i^{0,1}h_0(R_i^{0,1}))\right.$$
$$\left. + \frac{1}{4}\sum_{i=1}^{k-1}\mathrm{E}((R_i^{0,1})^3 h_0(R_i^{0,1})) - \frac{1}{4}\sum_{i=1}^{k-1}\mathrm{E}((R_i^{0,1})^2 h_0^2(R_i^{0,1}))\right],$$

$$I'_{33} = -\sum_{i=1}^{k}\mathrm{E}\left(\frac{\partial^2}{\partial\mu_2^2}L^*(\theta; R_i, R_{[i]})\right) = \frac{k}{\sigma_2^2(1-\rho^2)},$$

and

$$I'_{44} = -\sum_{i=1}^{k}\mathrm{E}\left(\frac{\partial^2}{\partial(\sigma_2^2)^2}L^*(\theta; R_i, R_{[i]})\right)$$
$$= \frac{1}{\sigma_2^4}\left[\frac{k}{2} + \frac{\rho^2}{4(1-\rho^2)}\sum_{i=1}^{k}\alpha_i^{(2)}\right].$$

Amini and Ahmadi (2007) showed that $I'_{55} = \frac{(1-\rho^2)\sum_{i=1}^{k}\alpha_i^{(2)} + 2k\rho^2}{(1-\rho^2)^2}$. Furthermore, we obtain similarly

$$I'_{12} = \frac{1}{\sigma_1^3}\left[\frac{1-\rho^2/2}{1-\rho^2}\sum_{i=1}^{k}\alpha_i - \frac{1}{2}\left\{\sum_{i=1}^{k-1}\mathrm{E}(h_0(R_i^{0,1})) - \sum_{i=1}^{k-1}\mathrm{E}((R_i^{0,1})^2 h_0(R_i^{0,1}))\right.\right.$$
$$\left.\left. + \sum_{i=1}^{k-1}\mathrm{E}(R_i^{0,1}h_0^2(R_i^{0,1}))\right\}\right], I'_{13} = \frac{-k\rho}{\sigma_1\sigma_2(1-\rho^2)}, \quad I'_{14} = \frac{-\rho^2}{2\sigma_1\sigma_2^2(1-\rho^2)}\sum_{i=1}^{k}\alpha_i,$$

$$I'_{15} = \frac{-\rho}{\sigma_1(1-\rho^2)}\sum_{i=1}^{k}\alpha_i, \quad I'_{23} = \frac{-\rho}{2\sigma_1^2\sigma_2(1-\rho^2)}\sum_{i=1}^{k}\alpha_i,$$

$$I'_{24} = \frac{-\rho^2}{4\sigma_1^2\sigma_2^2(1-\rho^2)}\sum_{i=1}^{k}\alpha_i^{(2)}, \quad I'_{25} = \frac{-\rho}{2\sigma_1^2(1-\rho^2)}\sum_{i=1}^{k}\alpha_i^{(2)},$$

$$I'_{34} = \frac{\rho}{2\sigma_2^3(1-\rho^2)}\sum_{i=1}^{k}\alpha_i, \quad I'_{35} = \frac{1}{\sigma_2(1-\rho^2)}\sum_{i=1}^{k}\alpha_i,$$

and

$$I'_{45} = \frac{\rho}{2\sigma_2^2(1-\rho^2)}\left[\sum_{i=1}^{k}\alpha_i^{(2)} - 2k\right].$$

The Fisher information matrix of the first k bivariate records and record times, $I_{RCT}(\theta)$, is obtained similarly.

5. Asymptotic relative efficiencies

The Fisher information matrices are free of μ_1 and μ_2. It can be seen that $I^{-1}(1,1)$ is multiple of σ_1^2, $I^{-1}(2,2)$ is multiple of σ_1^4, $I^{-1}(3,3)$ is multiple of σ_2^2, $I^{-1}(4,4)$ is multiple of σ_2^4 and $I^{-1}(5,5)$ is free of σ_1 and σ_2, for all three mentioned information matrices, where $A(i,i)$ is the element of the ith row and ith column of the matrix A.

Thus, the asymptotic relative efficiency (denoted by ARE) given below does not depend on μ_i and σ_i, $i = 1, 2$. Hence, without loss of generality, we take $\sigma_1 = \sigma_2 = 1$ in evaluating I_{RC}^{-1}, I_{RCT}^{-1} and I_{IID}^{-1} using the above formulas. Asymptotic efficiency of the MLE using bivariate record values and times with respect to (w.r.t.) that using bivariate random sample of the same size, and ARE of the MLE using bivariate record values and times w.r.t. that using bivariate record values without times are as follows:

$$\text{ARE}(\hat{\theta}_{i,RCT};\hat{\theta}_{i,IID}) = \frac{I_{IID}^{-1}(i,i)}{I_{RCT}^{-1}(i,i)};\ \ \text{ARE}(\hat{\theta}_{i,RCT};\hat{\theta}_{i,RC}) = \frac{I_{RC}^{-1}(i,i)}{I_{RCT}^{-1}(i,i)},\ i = 1,\ldots,5.$$

Tables 1 to 5 show the values of ARE of $\mu_1, \sigma_1, \mu_2, \sigma_2$ and ρ, respectively. These values are similar for negative and positive values of ρ. Since the values variations for Tables 1 and 2 w.r.t. ρ are very negligible, they are shown in 8 decimal places. The other tables' values are shown in 3 decimal places.

Table 1: The values of $\text{ARE}(\hat{\mu}_{1,RCT};\hat{\mu}_{1,IID})$ $(\text{ARE}(\hat{\mu}_{1,RCT};\hat{\mu}_{1,RC}))$.

		k	
ρ	2	3	4
0.1	0.75230738(1.17312564)	0.57576506(1.28298358)	0.49074455(1.43525201)
0.3	0.75230737(1.17312564)	0.57576512(1.28298355)	0.49074457(1.43525195)
0.5	0.75230713(1.17312566)	0.57576506(1.28298357)	0.49074461(1.43525236)
0.7	0.75230716(1.17312565)	0.57576490(1.28298304)	0.49074453(1.43525122)
0.9	0.75230687(1.17312497)	0.57576826(1.28298400)	0.49074460(1.43525451)

Table 2: The values of $\text{ARE}(\hat{\sigma}_{1,RCT};\hat{\sigma}_{1,IID})$ $(\text{ARE}(\hat{\sigma}_{1,RCT};\hat{\sigma}_{1,RC}))$.

		k	
ρ	2	3	4
0.1	1.17304229(1.26641666)	1.13989568(1.25907952)	1.18477633(1.31711383)
0.3	1.17304227(1.26641666)	1.13989568(1.25907947)	1.18477632(1.31711383)
0.5	1.17304236(1.26641667)	1.13989570(1.25907955)	1.18477626(1.31711405)
0.7	1.17304220(1.26641663)	1.13989544(1.25907899)	1.18477616(1.31711337)
0.9	1.17304198(1.26641643)	1.13989669(1.25908029)	1.18477632(1.31711383)

Table 3: The values of ARE($\hat{\mu}_{2,RCT};\hat{\mu}_{2,IID}$) (ARE($\hat{\mu}_{2,RCT};\hat{\mu}_{2,RC}$)).

			ρ		
k	0.1	0.3	0.5	0.7	0.9
2	0.842 (1.002)	0.834 (1.017)	0.818 (1.047)	0.796 (1.090)	0.768 (1.143)
3	0.652 (1.003)	0.645 (1.029)	0.632 (1.078)	0.613 (1.148)	0.589 (1.235)
4	0.526 (1.005)	0.523 (1.042)	0.5169(1.115)	0.508 (1.221)	0.497 (1.357)
5	0.444 (1.007)	0.446 (1.060)	0.450 (1.169)	0.457 (1.337)	0.466 (1.568)

Table 4: The values of ARE($\hat{\sigma}_{2,RCT};\hat{\sigma}_{2,IID}$) (ARE($\hat{\sigma}_{2,RCT};\hat{\sigma}_{2,RC}$)).

			ρ		
k	0.1	0.3	0.5	0.7	0.9
2	1.002 (1.00002)	1.016(1.0019)	1.043 (1.0148)	1.085 (1.059)	1.141 (1.170)
3	1.003 (1.00002)	1.030(1.0019)	1.078 (1.015)	1.130 (1.062)	1.154 (1.172)
4	1.005 (1.00003)	1.043(1.002)	1.114 (1.019)	1.191 (1.077)	1.217 (1.214)
5	1.006 (1.00003)	1.056(1.003)	1.153 (1.027)	1.280 (1.117)	1.371 (1.341)

Table 5: The values of ARE($\hat{\rho}_{RCT};\hat{\rho}_{IID}$) (ARE($\hat{\rho}_{RCT};\hat{\rho}_{RC}$)).

			ρ		
k	0.1	0.3	0.5	0.7	0.9
2	1.094(1.001)	1.093(1.011)	1.090(1.031)	1.087(1.061)	1.082(1.100)
3	1.206(1.001)	1.193(1.012)	1.168(1.033)	1.133(1.063)	1.090(1.100)
4	1.323(1.002)	1.300(1.016)	1.256(1.042)	1.196 (1.078)	1.124 (1.122)
5	1.444(1.003)	1.415(1.024)	1.361(1.065)	1.287(1.120)	1.200(1.184)

As we can see from the Tables 1 to 5, considering inter-record times along by bivariate records, provides a distinct improve in estimation of θ. Also, bivariate records and times lead to more precise estimators of scale and correlation parameters of bivariate normal distribution. On the other hand, location parameters' estimates are more precise based on IID database. Furthermore, greater correlation and smaller k leads to improvement of the estimators based on records relative to the estimators based on iid sample.

References

1. Amini, M. and Ahmadi, J., 2007. Fisher information in record values and their concomitants about the dependence and correlation parameters. *Statist. Probab. Lett.* **77**, 964 – 972.
2. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., 1998. *Records*, John Wiley & Sons, New York.

Confidence Intervals and Regions for the Two-Parameter Exponential Distribution Based on Records

A. Asgharzadeh and M. Abdi

Department of Statistics, University of Mazandaran

The exponential distribution is widely adopted as a lifetime model. Many authors have considered the interval estimation of the parameters of two-parameter exponential distribution based on complete and censored samples. In this paper, we consider the interval estimation of the location and scale parameters and the joint confidence region of the parameters of two-parameter exponential distribution based on upper records. A simulation study is done for the performance of all proposed confidence intervals and regions. We also propose the predictive intervals of the future records. Finally, a numerical example is given to illustrate the proposed methods.

Keywords: Confidence interval; Confidence region; Exponential distribution; Predictive interval; Record values.

1. Introduction

Let $X_1,\ X_2, ...$ be a sequence of independent and identically distributed (iid) random variables with cumulative distribution function (cdf) $F(x;\theta)$ and probability density function (pdf) $f(x;\theta)$, where θ is an unknown parameter. An observation X_j will be called an upper record value if its value exceeds that of all previous observations. Thus, X_j is an upper record if $X_j > X_i$ for every $i < j$. An analogous definition can be given for lower record values. If $\{U(n),\ n \geq 1\}$ is defined by

$$U(1) = 1, \qquad U(n) = \min\{j : j > U(n-1),\ X_j > X_{U(n-1)}\}, \tag{1}$$

for $n \geq 2$, then the sequence $\{X_{U(n)},\ n \geq 1\}$ provides a sequence of upper record statistics. The sequence $\{U(n),\ n \geq 1\}$ represents the record times. For more details and applications of record values, see Arnold et al. (1998).

The exponential distribution is applied in very wide variety of statistical procedures, especially in lifetime data analysis and reliability. A great deal of research has been done on statistical inference for the exponential distribution, and a very good summary of this work can be found in Johnson et. al (1994). The cumulative distribution function (cdf) of the two-parameter exponential distribution is given

by

$$F(x;\mu,\sigma) = 1 - \exp(-\frac{x-\mu}{\sigma}), \quad x \geq \mu, \quad \sigma > 0, \tag{2}$$

where μ and σ are the location and scale parameters, respectively.

Recently, some authors have considered the interval estimation of the parameters of two-parameter exponential distribution based on censored samples. Sun et. al. (2008) obtained confidence intervals for the scale parameter of exponential distribution based on Type-II doubly censored samples. Wu (2007, 2010) proposed the interval estimation of the scale parameter and the joint confidence region of the parameters of two-parameter exponential distribution based on doubly Type-II censored sample and progressively Type-II censored sample, respectively. But up to now, no work has been done on confidence intervals and regions of the parameters of two-parameter exponential distribution based on records. In this paper, two methods propose for constructing confidence intervals and joint confidence regions for the parameters of two-parameter exponential distribution based on records.

The paper is organized as follows: In Section 2, confidence intervals for μ and σ are considered based on upper records. We also present two exact joint confidence regions for the parameters μ and σ. In Section 3, a simulation study is performed to compare the proposed confidence intervals and regions. In addition to the estimation of the two parameters, the predictive intervals of the future record $X_{U(n+1)}$ based on observed upper records are given in Section 4. Finally in Section 5, a numerical examples is presented to illustrate the methods.

2. Confidence intervals and regions

Let $X_{U(1)} < X_{U(2)} < \cdots < X_{U(m)}$ be the first m observed upper record values from the two-parameter exponential distribution. For notation simplicity, we will write X_i for $X_{U(i)}$. Let $Y_i = (X_i - \mu)/\sigma$, $(i = 1, 2, \ldots, m)$, then $Y_1 < Y_2 < \cdots < Y_m$ are the first m upper record values from a standard exponential distribution. Moreover, $Z_1 = Y_1$ and $Z_i = Y_i - Y_{i-1}$, for $i = 2, ..., m$, are $i.i.d$ standard exponential random variables. Hence

$$U = 2\ Z_1 = 2\ Y_1, \tag{3}$$

has a chi-square distribution with 2 degrees of freedom and

$$V = 2\sum_{i=2}^{m} Z_i = 2\ (Y_m - Y_1), \tag{4}$$

has a chi-square distribution with $2m - 2$ degrees of freedom. We can also find that U and V are independent random variables. Now, let

$$T_1 = \frac{U/2}{V/2(m-1)} = \frac{(m-1)Y_1}{Y_m - Y_1}, \tag{5}$$

and

$$T_2 = U + V = 2\ Y_m. \tag{6}$$

It is easy to show that T_1 has an F distribution with 2 and $2m-2$ degrees of freedom and T_2 has a chi-square distribution with $2m$ degrees of freedom. Furthermore, T_1 and T_2 are independent (see Johnson et al. (1994), Page 350).

The distributions of the pivotal quantities U, V, T_1 and T_2 are independent of parameters. Make use of these pivotal quantities, we can construct the confidence intervals and regions for parameters. Using the pivotal quantity T_1, we can construct a confidence interval for the location parameter μ as follows:

Theorem 2.1. *Suppose that $X_1 < X_2 < \cdots < X_m$ be the first m observed upper record values from the two-parameter exponential distribution in (2). Then, for any $0 < \alpha < 1$, the $(1-\alpha)100\%$ confidence intervals of the location parameter μ is given by*

$$\left(X_1 - \frac{(X_m - X_1)F_{\frac{\alpha}{2}}(2, 2m-2)}{m-1} < \mu < X_1 - \frac{(X_m - X_1)F_{1-\frac{\alpha}{2}}(2, 2m-2)}{m-1}\right),$$

where $F_{\frac{\alpha}{2}}(2, 2m-2)$ is the right-tailed $\frac{\alpha}{2}$ percentile for F distribution with 2 and $2m-2$ degrees of freedom.

Proof. From (5), we know that

$$T_1 = \frac{(m-1)(X_1 - \mu)}{X_m - X_1} \sim F(2, 2m-2),$$

then we have

$$\begin{aligned} 1-\alpha &= P\left(F_{1-\frac{\alpha}{2}}(2, 2m-2) < T_1 < F_{\frac{\alpha}{2}(2,2m-2)}\right) \\ &= P\left(X_1 - \frac{(X_m - X_1)F_{\frac{\alpha}{2}}(2, 2m-2)}{m-1} < \mu < X_1 - \frac{(X_m - X_1)F_{1-\frac{\alpha}{2}}(2, 2m-2)}{m-1}\right). \end{aligned}$$

The proof is thus obtained. □

Make use of the pivotal quantity V, we can construct a confidence interval for the scale parameter σ as follows:

Theorem 2.2. *Based on the first m observed upper record values $X_1 < X_2 < \cdots < X_m$, a $(1-\alpha)100\%$ confidence interval for the scale parameter σ is given by*

$$\left(\frac{2(X_m - X_1)}{\chi^2_{\frac{\alpha}{2}}(2m-2)} < \sigma < \frac{2(X_m - X_1)}{\chi^2_{1-\frac{\alpha}{2}}(2m-2)}\right),$$

where $\chi^2_{\frac{\alpha}{2}}(2m-2)$ is the right-tailed $\frac{\gamma}{2}$ percentile for chi-square distribution with $2m-2$ degrees of freedom.

Proof. From (4), we note that

$$V = \frac{2(X_m - X_1)}{\sigma} \sim \chi^2(2m-2).$$

Hence, we have

$$\begin{aligned} 1-\alpha &= P\left(\chi^2_{1-\frac{\alpha}{2}}(2m-2) < V < \chi^2_{\frac{\alpha}{2}}(2m-2)\right) \\ &= P\left(\frac{2(X_m - X_1)}{\chi^2_{\frac{\alpha}{2}}(2m-2)} < \sigma < \frac{2(X_m - X_1)}{\chi^2_{1-\frac{\alpha}{2}}(2m-2)}\right). \end{aligned}$$

□

Using the set of pivotal quantities U and V, we can construct a confidence region for the parameters μ and σ in the following Theorem. We call it Method 1.

Theorem 2.3. *(**Method 1**) Based on the pivotal quantities U and V, a $(1-\alpha)100\%$ joint confidence region of two parameters μ and σ is given by*

$$\begin{cases} X_1 - \frac{\sigma}{2}\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2) < \mu < X_1 - \frac{\sigma}{2}\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2), \\ \\ \frac{2(X_m - X_1)}{\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m-2)} < \sigma < \frac{2(X_m - X_1)}{\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m-2)}. \end{cases}$$

Proof. Since U and V are independent and $U = 2(X_1 - \mu)/\sigma \sim \chi^2(2)$ and $V = 2(X_m - X_1)/\sigma \sim \chi^2(2m-2)$, then for any $0 < \alpha < 1$, we have

$$P\left(\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2) < U < \chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2)\right) = \sqrt{1-\alpha},$$

and

$$P\left(\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m-2) < V < \chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m-2)\right) = \sqrt{1-\alpha}.$$

From these relationships, we obtain

$$\begin{aligned} &P\left(\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2) < \frac{2(X_1-\mu)}{\sigma} < \chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2),\right. \\ &\left.\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m-2) < \frac{2(X_m - X_1)}{\sigma} < \chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m-2)\right) \\ &= \sqrt{1-\alpha}.\sqrt{1-\alpha} = 1-\alpha. \end{aligned}$$

Equivalently,

$$\begin{aligned} P\Big(&X_1 - \frac{\sigma}{2}\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2) < \mu < X_1 - \frac{\sigma}{2}\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2), \\ &\frac{2(X_m - X_1)}{\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m-2)} < \sigma < \frac{2(X_m - X_1)}{\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m-2)}\Big) = 1-\alpha. \end{aligned}$$

Thus the theorem follows. □

Now, make use of the set of pivotal quantities T_1 and T_2, we can construct another confidence region for the parameters of μ and σ in the following Theorem. We call it Method 2.

Theorem 2.4. ***(Method 2)*** *Based on the pivotal quantities T_1 and T_2, a $(1-\alpha)100\%$ joint confidence region for the parameters of μ and σ is given by*

$$\begin{cases} X_1 - \frac{(X_m - X_1)F_{\frac{1-\sqrt{1-\alpha}}{2}}(2,2m-2)}{m-1} < \mu < X_1 - \frac{(X_m - X_1)F_{\frac{1+\sqrt{1-\alpha}}{2}}(2,2m-2)}{m-1}, \\ \frac{2(X_m-\mu)}{\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m)} < \sigma < \frac{2(X_m-\mu)}{\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m)}. \end{cases}$$

Proof. From (5) and (6), we know that

$$T_1 = \frac{(m-1)(X_1-\mu)}{X_m - X_1},$$

has an F distribution with 2 and $2m-2$ degrees of freedom, and

$$T_2 = \frac{2}{\sigma}(X_n - \mu)$$

has a chi-square distribution with $2n$ degrees of freedom, and it is independent of T_1. Next, for $0 < \alpha < 1$, we have

$$P\left(F_{\frac{1+\sqrt{1-\alpha}}{2}}(2,2m-2) < T_1 < F_{\frac{1-\sqrt{1-\alpha}}{2}}(2,2m-2)\right) = \sqrt{1-\alpha},$$

and

$$P\left(\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m) < T_2 < \chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m)\right) = \sqrt{1-\alpha}.$$

From these relationships, we obtain

$$P\Big(F_{\frac{1+\sqrt{1-\alpha}}{2}}(2,2m-2) < T_1 < F_{\frac{1-\sqrt{1-\alpha}}{2}}(2,2m-2), \\ \chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m) < T_2 < \chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m)\Big) = 1-\alpha.$$

Then, the following inequalities determine $(1-\alpha)100\%$ joint confidence region of two parameters μ and σ,

$$\begin{cases} X_1 - \frac{(X_m - X_1)F_{\frac{1-\sqrt{1-\alpha}}{2}}(2,2m-2)}{m-1} < \mu < X_1 - \frac{(X_m - X_1)F_{\frac{1+\sqrt{1-\alpha}}{2}}(2,2m-2)}{m-1}, \\ \frac{2(X_m-\mu)}{\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m)} < \sigma < \frac{2(X_m-\mu)}{\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m)}, \end{cases}$$

The proof is thus completed. □

3. Simulation

In this section, we carry out a Monte Carlo simulation to illustrate all the methods discussed in this paper. The computations are performed using Visual S-PLUS package. In this simulation, we randomly generate upper record sample $X_1 < X_2 < \cdots < X_m$ from the standard exponential distribution with $\mu = 0$ and $\sigma = 1$, and then computed 90% confidence intervals and regions using Theorems 2.1-2.4. We replicate the process 5000 times. For given parameters $(\mu, \sigma) = (0, 1)$, the simulated average confidence length and confidence area with the confidence coefficient $1 - \alpha = 0.90$ for $m = 5(1)15$ are presented in Table 1. From Table 1, we observe when m increases, the average confidence lengths for μ and σ, and the average confidence area in both Methods, 1 and 2 are decreased. We also observe that Method 2 has better performance than method 1, since it provides the smaller confidence area.

Table 1. The simulated average confidence length and confidence area.

	Length		Area	
m	μ	σ	Method 1	Method 2
5	4.4208	2.4191	29.7751	26.9818
6	4.1192	2.0251	20.4781	18.7235
7	3.8209	1.7197	14.8141	13.6641
8	3.7085	1.5484	12.3008	11.4338
9	3.5660	1.3948	10.0895	9.4405
10	3.4964	1.2908	8.8476	8.3251
11	3.4299	1.2024	7.8700	7.4409
12	3.3870	1.1330	7.1541	6.7922
13	3.3577	1.0760	6.5934	6.2824
14	3.2830	1.0113	5.9769	5.7130
15	3.3110	0.9832	5.7680	5.5286

4. Prediction interval for the future record X_{m+1}

In this section, we present a prediction interval for the future record X_{m+1} based on the observed record values $X_1, ..., X_m$. Since the statistic

$$S = \frac{(m-1)(X_{m+1} - X_m)}{X_m - X_1},$$

follows a F distribution with 2 and $2m - 2$ degrees of freedom, we can construct a prediction interval for X_{m+1} as follows:

Theorem 4.1. *Based on the observed record values $X_1 < X_2 < \cdots < X_m$, a $(1-\alpha)100\%$ prediction interval for the future observation X_{m+1} is given by*

$$\left(X_m + \left(\frac{X_m - X_1}{m-1}\right) F_{1-\frac{\alpha}{2}}(2, 2m-2), X_m + \left(\frac{X_m - X_1}{m-1}\right) F_{\frac{\alpha}{2}}(2, 2m-2)\right).$$

Proof. Since $S \sim F(2, 2m-2)$, then we have

$$\begin{aligned} 1-\alpha &= P\left(F_{1-\frac{\alpha}{2}}(2, 2m-2) < \frac{(m-1)(X_{m+1} - X_m)}{X_m - X_1} < F_{\frac{\alpha}{2}}(2, 2m-2)\right) \\ &= P\left(X_m + \left(\frac{X_m - X_1}{m-1}\right) F_{1-\frac{\alpha}{2}}(2, 2m-2) < X_{m+1} < \right. \\ &\qquad \left. X_m + \left(\frac{X_m - X_1}{m-1}\right) F_{\frac{\alpha}{2}}(2, 2m-2)\right). \end{aligned}$$

The proof is thus completed. □

5. Numerical Example

In this example, we consider a simulated sample of size $n = 15$ from the Exponential distribution in (2) with parameters $\mu = 2.5$ and $\sigma = 1$. The simulated upper records are as follows:

```
2.7127  3.1841  3.4990  4.0002  4.4863  4.6226  5.3125  5.9537
5.9724  0.0238  6.0357  6.8307  6.8957  7.4296  8.0229.
```

By Theorem 2.1 and 2.2, the 95% confidence intervals for μ and σ are obtained as (1.44574,2.69325) and (0.25692,0.62739) with confidence lengths 1.24751 and 0.37047, respectively.

By Theorem 2.3 (Method 1), the 95% joint confidence region for μ and σ is given by

$$\begin{cases} 2.7127 - \frac{\sigma}{2} 8.73857 < \mu < 2.7127 - \frac{\sigma}{2} 0.02548, \\ \\ 0.22442 < \sigma < 0.75975, \end{cases}$$

with area 1.14763. Applying Theorem 2.4 (Method 2), the confidence region is given by

$$\begin{cases} 0.76774 < \mu < 2.70791, \\ \\ \frac{2(8.0229-\mu)}{49.9138} < \sigma < \frac{2(8.0229-\mu)}{15.3903}, \end{cases}$$

with area 1.09604. By Theorem 4.1, the prediction interval for X_{16} is obtained as (8.04239, 9.28991).

References

1. B.C. Arnold and N. Balakrishnan, H. N. Nagaraja, *Records*. New York, John Wiley and Sons, 1998.
2. N.L. Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distribution*. vol. 1. New York, John Wiley and Sons, 1994.
3. X. Sun, X. Zhou, J. Wang, Confidence intervals for the scale parameter of exponential distribution based on Type II doubly censored samples. *J. Stat. Plan. Inf.* **138**, 2045-2058, 2008.
4. S.F. Wu, Interval estimation for the two-Parameter Exponential distribution based on the doubly type II censored sample. *Quality & Quantity*, **41**, 489-496, 2007.
5. S.F. Wu, Interval estimation for the two-parameter exponential distribution under progressive censoring. *Quality & Quantity*, **44**, No.1,181-189, 2010.

Bivariate Maximum Entropy Models with Given Marginal Distributions

S. Ashrafi and M. Asadi

Department of Statistics, University of Isfahan

Let $\mathbf{X} = (X_1, X_2)$ be a continuous random vector. In this paper, under the assumption that the marginal distributions of X_1 and X_2 are given, we develop models for vector $\mathbf{X}$ when there is some partial information about the dependence structure between X_1 and X_2. The models, are obtained based on well-known Principle of Maximum Entropy, called the maximum entropy (ME) models. Our results lead to characterization of some well-known bivariate distributions such as Generalized Gumbel, Farlie-Gumbel-Morgenstern and Clayton bivariate distributions. The relationship between ME models and some well known dependence notions are studied. Conditions under which the mixture of bivariate distributions are ME models are also investigated.

Keywords: Fréchet class of distributions, Hazard gradient, Dependence, Total positive of order 2.

1. Introduction

Let $\mathbf{X} = (X_1, X_2)$ be a continuous random vector with distribution function $F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$, $x_1, x_2 \in R$. Specification of a bivariate distribution requires full information about marginal distributions as well as dependence structure between X_1 and X_2. There are many situations in which the marginal distributions are known but the complete information about the dependence structure between X_1 and X_2 is unknown. The problem of interest, in such situations, is to make inference about joint distribution based on constraints on some specifications of the population. A well known approach to characterize a model for the data generating distribution is the maximum entropy method. In this approach, insufficient knowledge about the data generating distribution is formulated in terms of a set of information constraints. Usually the constraints are made on the moments of the model and the aim is to find the model that maximizes Shannon entropy under these constraints.

In reliability engineering and survival analysis, there are several criterion which play the central role in analyzing the lifetime data in both univariate and multivariate case. Among the well known measures, the hazard rate and mean resid-

ual life function are of particular interest. The present paper provides a solution for problem of specification of bivariate models using the well-known Principle of Maximum Entropy; especially when partial information about the dependence structures between X_1 and X_2 are available and the constraints are made based on hazard gradient or reversed hazard gradient. In the univariate case, when the constraints are based on hazard rate and mean residual life function, Asadi et al [1] studied a concept of maximum dynamic entropy and Asadi et al [2] introduced a notion of minimum dynamic discrimination information and obtained various univariate lifetime distributions as maximum dynamic entropy and minimum dynamic discrimination models. Recently, Asadi et al [3] have studied a concept of bivariate dynamic ME model and derived several bivariate distributions when partial information is available on hazard gradient. This paper is a continuation of the work by Asadi et al [3]. In first part of this study, we develop several bivariate models based on partial information on the bivariate hazard gradient and reversed hazard gradient. This enables us to characterize several well-known bivariate distributions, with given marginal distributions, as maximum entropy (ME) models when partial information about hazard gradient or reversed hazard gradient are formulated in some inequality constraints. In the second part of the paper, we characterize mixture of bivariate distributions as ME models when partial information is available on the hazard rate of mixing distribution. The paper is organized as follows. Section 2 gives some preliminary results which are useful in subsequent sections. Section 3 gives results about ME models based on some dependence concepts such as total positivity of order 2 (TP_2) and reversed regularity of order 2 (RR_2) and some well known bivariate distributions are characterized as ME models under some partial information about the hazard and reversed hazard gradient. In Section 4 we discuss methods for characterizing bivariate mixtures as ME models when partial information about the distribution function of dependence parameter is available.

2. Preliminaries

in this section we give some definitions and preliminaries results which are useful in the subsequent sections. Let $\mathcal{M}(F_1, F_2)$ be the Fréchet class of absolutely continuous bivariate distributions functions (BDF) with given marginal distribution functions F_1 and F_2. The Kullback-Leibler discrimination information function between the BDF $F(x_1, x_2)$ and reference BDF $F_0(x_1, x_2)$ is defined by

$$K(F : F_0) = \int\int f(x_1, x_2) \log \frac{f(x_1, x_2)}{f_0(x_1, x_2)} dx_1 dx_2 \geq 0,$$

where F is absolutely continuous with respect to F_0 and f and f_0 denote the probability density functions (PDFs) of F and F_0, respectively. Note that $K(F :$

$F_0) = 0$ if and only if $f_0(x_1, x_2) = f(x_1, x_2)$ with probability 1. The joint entropy of F, denoted by $H(\mathbf{X})$, is defined to be

$$H(\mathbf{X}) = -\int\int f(x_1, x_2) \log f(x_1, x_2) dx_1 dx_2.$$

Definition 2.1. Let Ω_F be the set of all bivariate distributions in $\mathcal{M}(F_1, F_2)$ that satisfy some partial information (some constraints). The ME model in Ω_F is a DF $F^* \in \Omega_F$ such that

$$F^* = \arg \max_{F \in \Omega_F} H(F).$$

That is, the ME model is the one that its PDF maximizes joint entropy among all distributions in Ω_F [4].

In the univariate case, the hazard rate of a continuous distribution F with density f is defined as $\lambda(t) = \frac{f(t)}{\bar{F}(t)}$, $t > 0$, where $\bar{F}(t) = 1 - F(t)$. This function plays a main role in the study of lifetime random variables. In the literature there are various extensions of hazard rate $\lambda(t)$ to the multivariate. An extension is called the hazard gradient which is defined in the bivariate case as follows. Assume that the bivariate random vector $\mathbf{X}$ has the survival function $\bar{F}(x_1, x_2)$. The vector of hazard gradient of $\mathbf{X}$ is defined as

$$\begin{aligned}\Lambda_F(x_1, x_2) &= -(\frac{\partial \log \bar{F}(x_1, x_2)}{\partial x_1}, \frac{\partial \log \bar{F}(x_1, x_2)}{\partial x_2}) \\ &\equiv (\lambda_{F,1}(x_1, x_2), \lambda_{F,2}(x_1, x_2)).\end{aligned}$$

Note that $\lambda_{F,i}(x_1, x_2)$ is the hazard gradient of ith component. The hazard gradient $\Lambda_F(x_1, x_2)$ has the property that its relation to DF $F(x_1, x_2)$ is one-to-one. Another measure which is important in reliability and survival analysis is reversed hazard rate which is defined as $r(t) = \frac{f(t)}{F(t)}$. Similar to hazard gradient, in the bivariate case, the reversed hazard gradient is defined as

$$\begin{aligned}R_F(x_1, x_2) &= (\frac{\partial \log F(x_1, x_2)}{\partial x_1}, \frac{\partial \log F(x_1, x_2)}{\partial x_2}) \\ &\equiv (r_{F,1}(x_1, x_2), r_{F,2}(x_1, x_2)).\end{aligned}$$

Definition 2.2. Let the bivariate random vectors $\mathbf{X}$ and $\mathbf{Y}$ have distribution functions F and G which belong to $\mathcal{M}(F_1, F_2)$. $\mathbf{X}$ is said to be smaller than $\mathbf{Y}$ in positive quadrant dependent order (denoted by $\mathbf{X} \leq_{PQD} \mathbf{Y}$) if

$$\bar{F}(x_1, x_2) \leq \bar{G}(x_1, x_2) \quad \text{for all } (x_1, x_2).$$

A function $\phi : R^2 \longrightarrow R$ is said to be supermodular if for any $\mathbf{x}, \mathbf{y} \in R^2$

$$\phi(\mathbf{x}) + \phi(\mathbf{y}) \leq \phi(\mathbf{x} \wedge \mathbf{y}) + \phi(\mathbf{x} \vee \mathbf{y}),$$

where the operators $\wedge$ and $\vee$ denote coordinatewise minimum and maximum, respectively; see [5], p.335.

Definition 2.3. Let $\mathbf{X}$ and $\mathbf{Y}$ be two bivariate random vectors such that

$$E[\phi(\mathbf{X})] \leq E[\phi(\mathbf{Y})] \quad \text{for all supermodular functions } \phi : R^2 \longrightarrow R,$$

provided the expectations exist. Then $\mathbf{X}$ is said to be smaller than $\mathbf{Y}$ in supermodular order (denoted by $\mathbf{X} \leq_{sm} \mathbf{Y}$); see [5], p. 395.

Definition 2.4. (a) Let X_1 and X_2 have the joint density function $f(x_1, x_2)$. $f(x_1, x_2)$ is said to be TP_2 if

$$f(x_1, x_2)f(y_1, y_2) \geq f(x_1, y_2)f(y_1, x_2) \text{ for all } (x_1, x_2) < (y_1, y_2). \tag{1}$$

(b) X_2 is said to be right-tail increasing in X_1 (denoted by $RTI(X_2|X_1)$) if

$$P(X_2 > x_2|X_1 > x_1) \text{ increasing in } x_1 \text{ for all } x_2. \tag{2}$$

If the inequalities in (1) and (2) are reversed, the negative analogues can be obtained. Thus, the dual of (1) and (2) are, respectively, called reverse regular of order 2 (RR_2) and right tail decreasing (RTD).

3. ME models with given marginal distributions

In this section we study conditions under which a distribution is an ME model in Ω_F. The conditions are based on partial information about hazard gradient and reversed hazard gradient. Before giving the main results we need the following lemma.

Lemma 3.1. *Consider the bivariate random vectors* $\mathbf{X}$ *and* $\mathbf{Y}$ *with distribution functions* F *and* G *and hazard gradients* $\Lambda_F(\mathbf{x}) = (\lambda_{F,1}(\mathbf{x}), \lambda_{F,2}(\mathbf{x}))$ *and* $\Lambda_G(\mathbf{x}) = (\lambda_{G,1}(\mathbf{x}), \lambda_{G,2}(\mathbf{x}))$*, respectively. Assume that* $F, G \in \mathcal{M}(F_1, F_2)$*. If for all* $\mathbf{x}$*,* $\lambda_{F,i}(\mathbf{x}) \geq \lambda_{G,i}(\mathbf{x}),\ i = 1, 2$ *then* $\mathbf{X} \leq_{PQD} \mathbf{Y}$*.*

Proof: Let $\mathbf{X}$ and $\mathbf{Y}$ have survival functions $\bar{F}(\mathbf{x})$ and $\bar{G}(\mathbf{x})$. The condition that $\lambda_{F,i}(\mathbf{x}) \geq \lambda_{G,i}(\mathbf{x}),\ i = 1, 2$ is equivalent to $\frac{\bar{G}(\mathbf{x})}{\bar{F}(\mathbf{x})}$ is increasing in $\mathbf{x}$. Hence $\bar{G}(\mathbf{x}) \geq \bar{F}(\mathbf{x})$ for all $\mathbf{x}$. Using this and the assumption that $F, G \in \mathcal{M}(F_1, F_2)$, we have $\mathbf{X} \leq_{PQD} \mathbf{Y}$.

Remark 3.1. The result of Lemma 3.1 is also true if the marginal reversed hazard gradients are ordered. That is, under the condition that $F, G \in \mathcal{M}(F_1, F_2)$, if $r_{F,i}(\mathbf{x}) \geq r_{G,i}(\mathbf{x}),\ i = 1, 2$, then $\mathbf{X} \leq_{PQD} \mathbf{Y}$.

Now we can prove the following Theorem.

Theorem 3.1. *Let* $\Omega_F = \{F(x_1, x_2) \in \mathcal{M}(F_1, F_2) : \Lambda_F(x_1, x_2) \leq (\geq) Q(x_1, x_2)\}$ *be a set of distributions in* $\mathcal{M}(F_1, F_2)$ *having hazard gradient* Λ_F*. Suppose that there exists a distribution* $F^* \in \Omega_F$ *with PDF* f^* *such that* $\Lambda_{F^*}(x_1, x_2) = Q(x_1, x_2)$*. Then*

(a) F^* *is ME in* Ω_F *with* $\Lambda_F(x_1, x_2) \leq Q(x_1, x_2)$ *if* $f^*(x_1, x_2)$ *is* TP_2*.*
(b) F^* *is ME in* Ω_F *with* $\Lambda_F(x_1, x_2) \geq Q(x_1, x_2)$ *if* $f^*(x_1, x_2)$ *is* RR_2*.*

Proof: We prove part (a) of the theorem. Part (b) can be proved similarly. Let $\mathbf{X}$ and $\mathbf{Y}$ denote two bivariate random vectors with distribution functions F^* and F in $\mathcal{M}(F_1, F_2)$. Since $\Lambda_F(x_1, x_2) \leq \Lambda_{F^*}(x_1, x_2)$ for all $\mathbf{x}$ and F^* and F are in $\mathcal{M}(F_1, F_2)$, from Lemma 3.1, we have $\mathbf{X} \leq_{PQD} \mathbf{Y}$ which is equivalent to $\mathbf{X} \leq_{sm} \mathbf{Y}$ (see [5], p.395). Also it is easily seen that if $f^*(x_1, x_2)$ is TP_2 then $\log f^*(x_1, x_2)$ is supermodular. Thus

$$E[\log f^*(\mathbf{X})] \leq E[\log f^*(\mathbf{Y})] \tag{3}$$

On the other hand

$$K(F : F^*) = -H(\mathbf{Y}) - E[\log f^*(\mathbf{Y})] \geq 0.$$

Using this and (3) we obtain $H(\mathbf{X}) \geq H(\mathbf{Y})$. This completes the proof of the theorem .

Remark 3.2. It should be pointed that usually it may be difficult to check, using the definition, that a bivariate density $f^*(x_1, x_2)$ is TP_2 or RR_2. An alternative method is to use a result of Holland and Wang [6]. They showed that if $f^*(x_1, x_2)$ has second partial derivatives, then it is TP_2 (RR_2) if and only if $\frac{\partial^2}{\partial x_1 x_2} \log f^*(x_1, x_2) \geq (\leq) 0$.

Remark 3.3. Let f_1 and f_2 be marginal PDFs of distributions in $\mathcal{M}(F_1, F_2)$. Assume that $f_0(x_1, x_2) = f_1(x_1) f_2(x_2)$ is a reference distribution. Using Theorem 3.1 it can be proved that the ME model in Ω_F is a distribution that minimizes Kullback-Leibler discrimination information function with respect to reference distribution $f_0(x_1, x_2)$. In information literature, a criterion for measuring the dependency between F_1 and F_2 is the mutual information. If $M(X_1, X_2)$ denotes the mutual information between X_1 and X_2, then

$$\begin{aligned} M(X_1, X_2) &= H(X_1) - H(X_1|X_2) \\ &= K(F : F_1 F_2), \end{aligned}$$

where $H(X_1|X_2)$ denotes the entropy of conditional density $f(x_1|x_2)$. Using this, we conclude that the ME model in Ω_F is a distribution which has the minimum

dependency in the sense of mutual information. Also it is worth to note that if no information is available about dependency between F_1 and F_2, i.e. we assume $\Omega_F = \mathcal{M}(F_1, F_2)$, then the independent model $F(x_1, x_2) = F_1(x_1)F_2(x_2)$ is ME. This is so, because for all $F \in \Omega_F$ we have

$$K(F : F_1F_2) = -H(F) + H(F_1F_2) \geq 0$$

and equality occurs if and only if $F = F_1F_2$.

In the following we characterize some bivariate distributions as ME models where the constraints are based on hazard gradient. The key function on constructing the constraints is the ratio of ith element of hazard gradient $\Lambda(x_1, x_2) = (\lambda_1(x_1, x_2), \lambda_2(x_1, x_2))$ over the hazard rate of marginal distribution functions F_i, $i = 1, 2$. That is, to characterize the ME model we assume that partial information is available on $\eta_i(x_1, x_2)$, where

$$\eta_i(x_1, x_2) = \frac{\lambda_i(x_1, x_2)}{\lambda_i(x_i)}, \quad i = 1, 2. \tag{4}$$

The function $\eta_i(x_1, x_2)$ is closely related to well known concepts of dependency. In other words, assume that (X_1, X_2) has bivariate survival function $\bar{F}$. Then it is known that known that $\bar{F}$ is both $RTI(X_1|X_2)$ and $RTI(X_2|X_1)$ ($RTD(X_1|X_2)$ and $RTD(X_2|X_1)$) if and only if, for all x_1, x_2

$$\eta_i(x_1, x_2) \leq (\geq)1, \quad i = 1, 2. \tag{5}$$

Also it can shown that if $\bar{F}$ is TP_2 (RR_2) then (5) holds (see, [7]). This discussion leads to the following proposition.

Proposition 3.1. *Let* $\Omega_F = \{F(x_1, x_2) \in \mathcal{M}(F_1, F_2) : \Lambda_F(x_1, x_2) \leq (\geq) Q(x_1, x_2)\}$. *Assume that there exists a* $F^* \in \Omega_F$ *is such that* $\Lambda_{F^*}(x_1, x_2) = Q(x_1, x_2)$ *and* f^* *is* $TP_2(RR_2)$. *Then any* $F \in \Omega_F$ *is both* $RTI(X_1|X_2)$ *and* $RTI(X_2|X_1)$ *(*$RTD(X_1|X_2)$ *and* $RTD(X_2|X_1)$*). That is* $\eta_i(x_1, x_2) \leq (\geq)1, \quad i = 1, 2..$

Proof: Let $\Lambda_F(\mathbf{x}) = (\lambda_{F,1}(\mathbf{x}), \lambda_{F,2}(\mathbf{x}))$ be hazard gradient of a distribution F in Ω_F. First assume that f^* is TP_2. Then it can be shown that $\bar{F}^*$ is also TP_2. This implies that $\lambda_{F^*,i}(x_1, x_2) \leq \lambda_i(x_i)$ $i = 1, 2$. On the other hand, the assumption that $\lambda_{F,i}(x_1, x_2) \leq \lambda_{F^*,i}(x_1, x_2)$ gives $\lambda_{F,i}(x_1, x_2) \leq \lambda_i(x_i)$. Hence we have that F is both $RTI(X_1|X_2)$ and $RTI(X_2|X_1)$. When f^* is RR_2, the same arguments show that any $F \in \Omega$ is $RTD(X_1|X_2)$ and $RTD(X_2|X_1)$. This completes the proof.

The following theorem explores the relation between the covariance of the elements of ME model and the covariance of elements of any other distribution in Ω_F.

Theorem 3.2. *Let (X_1, X_2) and (Y_1, Y_2) be two bivariate vectors with distribution functions in $\mathcal{M}(F_1, F_2)$, and PDFs $f^*(\mathbf{x})$ and $f(\mathbf{x})$ and hazard gradients $\Lambda_{F^*}(\mathbf{x}) = (\lambda_{F^*,1}(\mathbf{x}), \lambda_{F^*,2}(\mathbf{x}))$ and $\Lambda_F(\mathbf{x}) = (\lambda_{F,1}(\mathbf{x}), \lambda_{F,2}(\mathbf{x}))$,respectively.*

(a) If $f^*(\mathbf{x})$ is TP_2 and $\Lambda_{F^*}(\mathbf{x}) \geq \Lambda_F(\mathbf{x})$ then $0 \leq Cov(X_1, X_2) \leq Cov(Y_1, Y_2)$.
(b) If $f^*(\mathbf{x})$ is RR_2 and $\Lambda_{F^*}(\mathbf{x}) \leq \Lambda_F(\mathbf{x})$ then $Cov(Y_1, Y_2) \leq Cov(X_1, X_2) \leq 0$.

Proof: We prove part (a). The proof of (b) is similar. It is known that when $f^*(\mathbf{x})$ is TP_2 then $Cov(X_1, X_2) \geq 0$ (see [8], p. 267). Since (X_1, X_2) and (Y_1, Y_2) have distribution functions in $\mathcal{M}(F_1, F_2)$, from Lemma 3.1, if $\Lambda_{F^*}(\mathbf{x}) \geq \Lambda_F(\mathbf{x})$ then $(X_1, X_2) \leq_{PQD} (Y_1, Y_2)$ which implies $Cov(X_1, X_2) \leq Cov(Y_1, Y_2)$. This completes the proof.

The result of this theorem shows that when $Cov(X_1, X_2)$ and $Cov(Y_1, Y_2)$ are covariances between the elements of ME model and elements of any other distribution in Ω_F, respectively, if $f^*(\mathbf{x})$ is ME in $\Omega_F = \{F(\mathbf{x}) \in \mathcal{M}(F_1, F_2) : \Lambda_F(\mathbf{x}) \leq (\geq)\Lambda_{F^*}(\mathbf{x})\}$ then $|Cov(X_1, X_2)| \leq |Cov(Y_1, Y_2)|$. This means that X_1 and X_2, the elements of ME model, have the minimum absolute value of the linear dependency in the class.

Now we have ready to give some examples.

Example 3.1. Let Ω_F be a subset of $\mathcal{M}(F_1, F_2)$ consisting all bivariate distributions with hazard gradient $\Lambda(x_1, x_2) = (\lambda_1(x_1, x_2), \lambda_2(x_1, x_2))$ that satisfies the following inequalities

$$\eta_i(x_1, x_2) \geq (1 - \delta \log \bar{F}_j(x_j)), \quad 0 \leq \delta \leq 1, \ i, j = 1, 2, \quad i \neq j, \tag{6}$$

where $\eta_i(x_1, x_2)$ is defined as (4). Then, the ME model in Ω_F is Generalized Gumbel distribution with PDF

$$\begin{aligned} f^*(x_1, x_2) = f_1(x_1) f_2(x_2)[(1 - \delta \log \bar{F}_1(x_1))(1 - \delta \log \bar{F}_2(x_2)) - \delta] \\ \times \exp\{-\delta \log \bar{F}_1(x_1) \log \bar{F}_2(x_2)\}, \end{aligned} \tag{7}$$

in which $0 \leq \delta \leq 1$ and f_1 and f_2 are marginal PDFs of distributions in $\mathcal{M}(F_1, F_2)$ and $\bar{F}_1, \bar{F}_2$ are survival functions associated to f_1, f_2. The validity of this result follows from part (b) of Theorem 3.1. To see this, one can easily show that

$$\frac{\partial^2}{\partial x_1 \partial x_2} log f^*(x_1, x_2) \leq 0$$

which is equivalent to say that f^* is RR_2. Also it is easy to see that $\lambda_i(x_i)(1 - \delta \log \bar{F}_j(x_j))$ is the ith element of hazard gradient of $f^*(x_1, x_2)$.

Example 3.2. Sarmanov [9] introduced a family of bivariate densities of the form

$$f^*(x_1, x_2) = f_1(x_1) f_2(x_2)\{1 + \omega \phi_1(x_1) \phi_2(x_2)\}, \quad x_1, x_2 \in R \tag{8}$$

where ω, ϕ_1 and ϕ_2 satisfy the following conditions

$$\int_{-\infty}^{\infty} \phi_i(u) f_i(u) du = 0, \ i = 1, 2 \ \text{ and } \ 1 + \omega\phi_1(x_1)\phi_2(x_2) \geq 0 \text{ for all } x_1, x_2. \quad (9)$$

It can be shown that $f^*(x_1, x_2)$ is $TP_2(RR_2)$ if

$$\omega\phi_1'(x_1)\phi_2'(x_2) \geq (\leq) 0 \ \text{ for all } x_1, x_2.$$

Let Ω_F be a set of bivariate distributions with marginal PDFs f_1 and f_2 whose hazard gradient $\Lambda(x_1, x_2) = (\lambda_1(x_1, x_2), \lambda_2(x_1, x_2))$ satisfies the following inequalities

$$\eta_i(x_1, x_2) \leq (\geq) \frac{1 + \omega\phi_i(x_i)\psi_j(x_j)}{1 + \omega\psi_1(x_1)\psi_2(x_2)}, \quad i, j = 1, 2, i \neq j \quad (10)$$

where

$$\psi_i(x) = \frac{1}{\bar{F}_i(x)} \int_x^{\infty} f_i(u)\phi_i(u) du, \ \ i = 1, 2$$

and ω, ϕ_1, ϕ_2 satisfy (9) with $\omega\phi_1'(x_1)\phi_2'(x_2) \geq (\leq) 0$. Using Theorem 3.1, f^* is ME in Ω_F since the ith element of hazard gradient of f^* is equal to

$$\lambda_i(x) \frac{1 + \omega\phi_i(x_i)\psi_j(x_j)}{1 + \omega\psi_1(x_1)\psi_2(x_2)}, \quad i, j = 1, 2, i \neq j$$

where $\lambda_i(x), i = 1, 2$ is hazard function with respect to $f_i(x)$.

There are members of family (8) for which the condition $\omega\phi_1'(x_1)\phi_2'(x_2) \geq (\leq) 0$ holds. One the most important member of the family is the Farlie-Gumbel-Morgenstern (FGM) bivariate distribution with PDF

$$f^*(x_1, x_2) = f_1(x_1) f_2(x_2)[1 + \alpha(1 - 2F_1(x_1))(1 - 2F_2(x_2))], \quad -1 \leq \alpha \leq 1, (11)$$

is a well known family of distributions with applications in various branches of statistics. For FGM model it can be shown that

$$\frac{\partial^2}{\partial x_1 \partial x_2} log f^*(x_1, x_2) \geq 0 \ (\leq 0) \quad \text{if} \ \ 0 \leq \alpha \leq 1 \ (-1 \leq \alpha \leq 0).$$

In other words f^* is $TP_2(RR_2)$ if $0 \leq \alpha \leq 1 (-1 \leq \alpha \leq 0)$. Let Ω_F be a set of bivariate distributions with marginal PDFs $f_1(x)$ and $f_2(x)$ and hazard gradient $\Lambda(x_1, x_2) = (\lambda_1(x_1, x_2), \lambda_2(x_1, x_2))$ that satisfies the following inequalities

$$\eta_i(x_1, x_2) \leq (\geq)[1 + \alpha F_j(x_j)(2F_i(x_i) - 1)][1 + \alpha F_1(x_1)F_2(x_2)]^{-1}, \quad (12)$$
$$0 \leq \alpha \leq 1 (-1 \leq \alpha \leq 0), \ i \neq j, \ i, j = 1, 2.$$

Under this constraint and the fact that $f^*(x_1, x_2)$ is $TP_2(RR_2)$ we get, from Theorem 3.1, that $f^*(x_1, x_2)$ is ME in Ω_F.

In some distributions constraints based on reversed hazard gradient are more simple than hazard gradient. In the following we give an example of this kind. The following theorem gives ME models in $\mathcal{M}(F_1, F_2)$, in which the constraints are made based on reversed hazard gradient. The proof of the theorem, which is similar to Theorem 3.1, is based on Remark 3.1 and hence is omitted.

Theorem 3.3. *Let* $\Omega_F = \{F(x_1, x_2) \in \mathcal{M}(F_1, F_2) : R_F(x_1, x_2) \geq (\leq) R(x_1, x_2)\}$ *be a set of distributions in* $\mathcal{M}(F_1, F_2)$ *having reversed hazard gradient* R_F*. Suppose that there exists a distribution* $F^* \in \Omega_F$ *with PDF* f^**such that* $R_{F^*}(x_1, x_2) = R(x_1, x_2)$*. Then*

(a) F^* *is ME in* Ω_F *with* $R_F(x_1, x_2) \leq R(x_1, x_2)$ *if* $f^*(x_1, x_2)$ *is* TP_2*.*
(b) F^* *is ME in* Ω_F *with* $R_F(x_1, x_2) \geq R(x_1, x_2)$ *if* $f^*(x_1, x_2)$ *is* RR_2*.*

Example 3.3. Consider the Clayton's bivariate distribution with PDF

$$f^*(x_1, x_2) = \frac{(\theta+1) f_1(x_1) f_2(x_2) (F_1(x_1) F_2(x_2))^{-\theta-1}}{[(F_1(x_1))^{-\theta} + (F_2(x_2))^{-\theta} - 1]^{\frac{1}{\theta}+2}}, \quad \theta > 0,$$

in which f_1 and f_2 are marginal densities with distribution functions F_1 and F_2, respectively. For this distribution it can be shown that

$$\frac{\partial^2}{\partial x_1 \partial x_2} log f^*(x_1, x_2) \geq 0.$$

Let Ω_F be a set of bivariate distributions with marginal distribution functions F_1 and F_2 and reversed hazard gradient $R(x_1, x_2) = (r_1(x_1, x_2), r_2(x_1, x_2))$ satisfying the following inequalities

$$\beta_i(x_1, x_2) \leq \frac{(F_i(x_i))^{-\theta}}{[(F_1(x_1))^{-\theta} + (F_2(x_2))^{-\theta} - 1]}, \quad , \theta > 0, \ i = 1, 2,$$

where $\beta_i(x_1, x_2) = \frac{r_i(x_1, x_2)}{r_i(x_i)}, i = 1, 2$ and $r_i(x), i = 1, 2$ is reversed hazard function of $F_i(x)$. Then the Clayton's bivariate distribution is ME in Ω_F. This follows from part (a) of Theorem 3.3 and the fact that $r_i(x_i)(F_i(x_i))^{-\theta}[(F_1(x_1))^{-\theta} + (F_2(x_2))^{-\theta} - 1]^{-1}$ is ith element of reversed hazard gradient of Clayton's bivariate distribution.

4. ME Models for Mixtures

Let $\mathcal{G} = \{G_\theta(x_1, x_2) \in \mathcal{M}(G_1, G_2), \theta \in \chi\}$ be a family of bivariate distribution functions, where χ is a subset of the real line and θ is dependence parameter between G_1 and G_2. In the Bayesian context the parameter θ is assumed to the realization of a random variable Θ with support in χ. Assuming that Θ has distribution function H, then H is known as the prior distribution. The mixture

of $\mathcal{G}$ with respect to prior distribution H, which is also known as the predictive distribution function, is defined as

$$F(x_1, x_2) = \int_{\chi} G_\theta(x_1, x_2) dH(\theta), \quad (x_1, x_2) \in R^2. \tag{13}$$

In this section we study the ME models in class of predictive models in bivariate setup for which the constraints are made on hazard rate of prior distribution $H(\theta)$. The key result is given in the following theorem.

Theorem 4.1. *Let $\Omega_H = \{H : \lambda_H(\theta) \leq (\geq) q(\theta)\}$ be the set of prior distributions with support in χ and hazard function $\lambda_H(\theta)$. Consider Ω_F as a set of predictive distributions of the form (13) in which $H(\theta) \in \Omega_H$ and $G_\theta(x_1, x_2) \in \mathcal{G}$ is given. Suppose that there exists a prior distribution $H^* \in \Omega_H$ such that $\lambda_{H^*}(\theta) = q(\theta)$ and let F^* with PDF f^* be predictive distribution with respect to H^*. If elements of hazard gradient G_θ are decreasing in θ then*

(a) F^* is ME in Ω_F relative to Ω_H with $\lambda_H(\theta) \leq q(\theta)$ if $f^*(x_1, x_2)$ is TP_2.
(a) F^* is ME in Ω_F relative to Ω_H with $\lambda_H(\theta) \geq q(\theta)$ if $f^*(x_1, x_2)$ is RR_2.

Proof: We prove part (a) of the theorem. The proof of (b) is similar. Let $F(x_1, x_2)$ and $F^*(x_1, x_2)$ be of the form (13) in which prior distributions have hazard functions $\lambda_H(\theta)$ and $\lambda_{H^*}(\theta)$, respectively. If $\lambda_H(\theta) \leq \lambda_{H^*}(\theta)$ and elements of hazard gradient G_θ are decreasing in θ then from Theorem 6.D.5 of [5],

$$\Lambda_F(x_1, x_2) \leq \Lambda_{F^*}(x_1, x_2), \ for\ all\ (x_1, x_2),$$

in which $\Lambda_F(x_1, x_2)$ and $\Lambda_{F^*}(x_1, x_2)$ are hazard gradient of F and F^*, respectively. Now we show that $F(x_1, x_2) \in \mathcal{M}(G_1, G_2)$. Let $f_1(x)$ and $f_2(x)$ be marginal PDFs of $F(x_1, x_2)$ and $g_i(x)$ be PDF of $G_i(x),\ i = 1, 2$. Then using Fubini's Theorem, we have

$$f_i(x_i) = \int\int g_\theta(x_1, x_2) dH(\theta) dx_j = \int\int g_\theta(x_1, x_2) dx_j dH(\theta) = g_i(x_i),$$
$$i \neq j, i, j = 1, 2.$$

If $f^*(x_1, x_2)$ is TP_2 then using part (a) of Theorem 3.1 the proof is complete.

Applications of Theorem 4.1 are given in the following example.

Example 4.1.

(a) Let Ω_H be the set of prior distributions with support $(0, 1)$ having hazard function $\lambda_H(\theta)$ such that

$$\lambda_H(\theta) \leq \frac{1}{1-\theta}, \quad \theta \in (0, 1).$$

Consider F-G-M bivariate distribution with $\alpha = \theta$ and $0 < \alpha < 1$. Let Ω_F be the class of mixtures of F-G-M bivariate distributions with respect to prior distributions $H(\theta) \in \Omega_H$. The uniform mixture of F-G-M distributions, which is F-G-M with $\alpha = \frac{1}{2}$ is ME in Ω_F. The result follows on noting that $\frac{1}{1-\theta}$ is the hazard function of uniform distribution on $(0, 1)$. Also it can be seen easily that elements of hazard gradient of F-G-M distribution are decreasing in α and that the PDF of F-G-M distribution is TP_2 for $\alpha = \frac{1}{2}$. Thus part (a) of Theorem 4.1 gives the result.

(b) Using the same arguments as used to prove part (a), assuming that $\Omega_H = \{H(\theta) : \lambda_H(\theta) \geq \frac{1}{1-\theta}, \quad \theta \in (-1, 0)\}$, part (b) of Theorem 4.1, implies that F-G-M distribution with $\alpha = \frac{-1}{2}$ is ME in Ω_F.

References

1. M. Asadi, N. Ebrahimi, G. G. Hamedani, and E. S. Soofi. Maximum dynamic entropy models. *J. Appl. Prob.*, **41**, (2004) 379-390.
2. M. Asadi, N. Ebrahimi, G. G. Hamedani, and E. S. Soofi. Minimum dynamic discrimination information models. *J. Appl. Prob.*, **42**, (2005) 643-660.
3. M. Asadi, S. Ashrafi, N. Ebrahimi, and E. S. Soofi. Models based on partial information about hazard gradient. To appear in *Prob. Eng. & Info. Sci.*
4. E. T. Jaynes. Information theory and statistical mechanics. *Physics Rev.*, **106**, (1957), 620-630.
5. M. Shaked, and J. G. Shanthikumar. *Stochastic Orders.* (Springer, 2007).
6. P. W. Holland, Y. J. Wang. Dependence function for continuous bivariate densities. *Communications in StatisticsTheory and Methods*, **16**, (1987), 863-876.
7. B. Khaledi, S. Kochar. Dependence, dispersiveness, and multivariate hazard rate ordering. *Probability in the Engineering and Information Sciences*, **19**, (2005), 427-446.
8. C. D. Lai, and M. Xie. *Stochastic Ageing and Dependence for Reliability.* (Springer, 2006)
9. O. V. Sarmanov. Generalized normal correlation and two-dimensional Frechét classes, *Doklady (Soviet Mathematics)*, **168**, (1966), 596-599.

On $M/(G_1, G_2)/1/G(BS)/V_s$ with Feedback on Each Service

A. Badamchizadeh

Department of Statistics, Allameh Tabataba'i University

In this system we consider an $M/G/1$ queue with two phases of heterogeneous services, Bernoulli feedback on each service and Bernoulli vacation , where the arrivals is Poisson. After the completion of the first phase with FCFS schedule , with probability p_1 the customer feedback to the tail of original queue or with probability $q_1 = 1-p_1$ the second phase of service starts.Again after completion of the second phase, with probability p_2 the customer's second phase of service cannot be accepted and he/she must repeat only this phase.With probability $q_2 = 1-p_2$ the services are completed and tagged customer departs the system. After completion of the second phase, the server takes a vacation with probability θ or may continue to stay in the system with probability $1-\theta$. The vacation times are assumed to be general.

Keywords: M/G/1 Queue,Two phase of heterogeneous service, Bernoulli feedback, Bernoulli vacation, Mean queue size, Mean response time.

1. Introduction

For the first time Keilson and Servi [5,6] studied the concept of Bernoulli vacation. They introduced the concept of modified service time which has a main rule in systems with general service and vacation .Considerable efforts have been devoted to this model by Servi[11], Ramaswam and Servi[9],Takagi[13], Doshi[4], Choi and Kim [2].

Choudhury and Paul[3] inspected the $M/G/1$ system with two phases of heterogeneous service and Bernoulli feedback. In this system a tagged customer may have an unsuccessful service, then retried until a successful service is completed.He/She with probability $p(0 \leqslant p \leqslant 1)$ is feedbacked instantaneously to the tail of the queue or departs from the system with probability $q = 1 - p$.

Also Madan and Choudhury [7] studied a single server queue with two phases of heterogeneous services under Bernoulli schedule and a general vacation time. In this system, without feedback , the server after the completion of the service, may go to vacation with probability θ or remain in system with probability $1-\theta$. Shahkar and Badamchi [12] performed this system with k phase of heterogeneous services.

Recently author works on similar models in [1] and [10].

In many examples such as cars check up systems, production systems, bank services, computer and communication networks, feedback occurs on each service and customers repeat only unsuccessful phase, also he/she feedbacks to the beginning of the this phase.In addition, for overhauling or maintenance of the system, the server may go to vacation.

In this paper we analyze a single server queue with Poisson input, two phases of heterogeneous service with Bernoulli feedback on each phase and Bernoulli vacation. In section 2 we deal with the mathematical model and definitions. Steady-State conditions and generating functions are obtained in section 3.

Mean queue size, mean response time and other measures of system are computed in section 4. Also, in section 5 we obtained results for some special cases which have already been worked and are mentioned in the above references.

2. Mathematical model and definitions

We consider a queueing system such that:
i) Customers arrive at the system one by one in a Poisson stream with mean rate $\lambda(> 0)$.
ii)The server provides two phases of heterogeneous service in succession. The service discipline is assumed to be first come, first served(FCFS). The service times for two phases are independent random variable; we denote them by B_1, B_2 and $B_1^*(s), B_2^*(s)$ for their Laplace-Stieltjes transform (LST); we assume they have finite moments $E(B_i^l)$ for $l \geqq 1$ and i=1,2.
iii)After the completion of first phase of service(FPS), if the unit is dissatisfied with its service for certain reason or if it received unsuccessful service, the customer may immediately join the tail of the original queue as a feedback customer for receiving another regular service with probability $p_1(0 \leqslant p_1 \leqslant 1)$.Otherwise the customer will be led to second phase of service(SPS) with probability $q_1 = 1 - p_1$.The probability of the second phase of service that is completed successfully is q_2 and then the customer departs the system, but with probability $p_2 = 1 - q_2$ this phase must be repeated, so the customer lead to the queue of the second phase.The service discipline for feedback and new customers are FCFS. Also the service time for a feedback customer is independent of its previous service times.

iv) As soon as the second phase of service of a customer is completed, the server may go for a vacation of random length V with probability $\theta(0 \leqq \theta \leqq 1)$ or it may continue to serve the next customer, if any, with probability $(1 - \theta)$; otherwise, it remains in the system and waits for a new arrival. We denote $V(x), V^*(s)$ and $E(V^l)$ for distribution function(DF), LST and l'th finite moment of V respectively where $l \geqq 1$.

Definition 2.1. The modified service time or the time required by a customer to complete the service cycle is given by

$$B = \begin{cases} B_1 + B_2 + V & \text{with probability } \theta \\ B_1 + B_2 & \text{with probability } (1-\theta) \end{cases} \tag{1}$$

then the LST $B^*(s)$ of B is given by

$$B^*(s) = \theta B_1^*(s) B_2^*(s) V^*(s) + (1-\theta) B_1^*(s) B_2^*(s) \tag{2}$$

and

$$E(B) = E(B_1) + E(B_2) + \theta\ E(V) \tag{3}$$

also

$$E(B^2) = E(B_1^2) + E(B_2^2) + 2E(B_1)E(B_2) + 2\theta E(V)[E(B_1) + E(B_2)] + \theta E(V^2) \tag{4}$$

Definition 2.2. The elapsed time of i-th phase of service $[(PS)_i]$at time 't' is denoted by $B_i^0(t)$ for $i = 1, 2$. Also $V^0(t)$ denotes the elapsed vacation time at time 't', and $N_Q(t)$ denotes the queue size at time 't' .For $i = 1, 2$ we introduce the random variable Y(t) as follow :

$$Y(t) = \begin{cases} 0 & \text{if the server is idle at time 't',} \\ i & \text{if the server is busy with } (PS)_i \text{ at time 't',} \\ 3 & \text{if the server is on vacation at time 't'.} \end{cases} \tag{5}$$

Then we have a bivariate Markov process $\{N_Q(t), L(t)\}$ where $L(t) = 0$ if $Y(t) = 0$; $L(t) = B_i^0(t)$ if $Y(t) = i$ for $i = 1, 2$ and $L(t) = V^0(t)$ if $Y(t) = 3$ .Now for $i = 1, 2$ the following probabilities are defined as

$$Q_n(x,t) = Prob[N_Q(t) = n, L(t) = V^0(t); x < V^0(t) \leqslant x + dx] \quad x > 0, \quad n \geqslant 0 \tag{6}$$

$$P_{i,n}(x,t) = Prob[N_Q(t) = n, L(t) = B_i^0(t); x < B_i^0(t) \leqslant x + dx] \quad x > 0, \quad n \geqslant 0 \tag{7}$$

and

$$R_0(t) = Prob[N_Q(t) = 0, L(t) = 0] \tag{8}$$

Assume that

$$V(0) = 0, V(\infty) = 1 \tag{9}$$

and for $i = 1, 2$

$$B_i(0) = 0, B_i(\infty) = 1 \tag{10}$$

Also $V(x)$ and $B_i(x)$ are continuous at x=0. Then we have the hazard rate functions of V and $B_i(i = 1, 2, ..., k)$ as follow

$$\nu(x) = \frac{dV(x)}{1 - V(x)} \tag{11}$$

$$\mu_i(x) = \frac{dB_i(x)}{1 - B_i(x)} \tag{12}$$

($\mu_i(x)$ be the conditional probability of the completion of $i - th$ stage of service during the time interval (x,x+dx), given that the elapsed service time is x.)

With the assumption that steady state exists, we let

$$R_0 = \lim_{t\to\infty} R_0(t) \tag{13}$$

$$P_{i,n}(x)dx = \lim_{t\to\infty} P_{i,n}(x, t)dx \quad i = 1, 2 \quad x > 0,\ n \geqslant 0 \tag{14}$$

$$Q_n(x)dx = \lim_{t\to\infty} Q_n(x, t)dx \quad x > 0, \quad n > 0 \tag{15}$$

Now for $i = 1, 2$ the PGF of this probabilities is defined as follows:

$$P_i(x, z) = \sum_{n=0}^{\infty} z^n P_{i,n}(x) \quad |z| \leqslant 1, \quad x > 0 \tag{16}$$

$$P_i(0, z) = \sum_{n=0}^{\infty} z^n P_{i,n}(0) \quad |z| \leqslant 1 \tag{17}$$

Also

$$Q(x, z) = \sum_{n=0}^{\infty} z^n Q_n(x) \quad |z| \leqslant 1, \quad x > 0 \tag{18}$$

$$Q(0, z) = \sum_{n=0}^{\infty} z^n Q_n(0) \tag{19}$$

3. Steady-state probability generating function

From Kolmogorov forward equations, for $i = 1, 2$ the steady-state conditions can be written as follows

$$\frac{d}{dx}P_{i,n}(x) + [\lambda + \mu_i(x)]P_{i,n}(x) = \lambda P_{i,n-1}(x) \quad n \geq 0, \quad x > 0 \tag{20}$$

and

$$\frac{d}{dx}Q_n(x) + [\lambda + \nu(x)]Q_n(x) = \lambda Q_{n-1}(x) \quad n \geq 0, \quad x > 0 \tag{21}$$

also

$$\lambda R_0 = \int_0^{+\infty} \nu(x)Q_0(x)dx + (1-\theta)q_2 \int_0^{+\infty} \mu_2(x)P_{2,0}(x)dx \tag{22}$$

We set $P_{1,-1}(x) = 0 \quad , P_{2,-1}(x) = 0$ and $Q_{-1}(x) = 0$ in (20), (21) and (22) .

At $x = 0$, the boundary conditions are

$$P_{1,0}(0) = \lambda R_0 + \{p_1 \int_0^{+\infty} \mu_1(x)P_{1,0}(x)dx + (1-\theta)q_2 \int_0^{+\infty} \mu_2(x)P_{2,1}(x)dx\} + \int_0^{+\infty} \nu(x)Q_1(x)dx \tag{23}$$

and for $n > 0$

$$P_{1,n}(0) = \{p_1 \int_0^{+\infty} \mu_1(x)P_{1,n}(x)dx + (1-\theta)q_2 \int_0^{+\infty} \mu_2(x)P_{2,n+1}(x)dx\} + \int_0^{+\infty} \nu(x)Q_{n+1}(x)dx \tag{24}$$

$$P_{2,n}(0) = p_2 \int_0^{+\infty} \mu_2(x)P_{2,n}(x)dx + q_1 \int_0^{+\infty} \mu_1(x)P_{1,n}(x)dx, \quad n \geqslant 0 \tag{25}$$

also

$$Q_n(0) = \theta q_2 \int_0^{+\infty} \mu_2(x)P_{2,n}(x)dx, \quad n \geqslant 0 \tag{26}$$

Finally the normalizing condition is

$$R_0 + \sum_{i=1}^{2}\sum_{n=0}^{\infty}\int_o^{+\infty} P_{i,n}(x)dx + \sum_{n=0}^{\infty}\int_0^{+\infty} Q_n(x)dx = 1 \tag{27}$$

Lemma 3.1. *For $i = 1, 2$ from (20) we have*

$$P_i(x,z) = P_i(0,z)[1 - B_i(x)]e^{-\lambda(1-z)x} \quad x > 0 \tag{28}$$

and from (21)

$$Q(x,z) = Q(0,z)[1 - V(x)]e^{-\lambda(1-z)x} \quad x > 0 \tag{29}$$

Proposition 3.1. *If for i=1,2*

$$B_i^*(\lambda-\lambda z)=\int_0^{+\infty} e^{-\lambda(1-z)x}dB_i(x) \tag{30}$$

$$V^*(\lambda-\lambda z)=\int_0^{+\infty} e^{-\lambda(1-z)x}dV(x) \tag{31}$$

are the z-transform of B_i and V respectively ,then

$$I)P_i(z)=P_i(0,z)[1-B_i^*(\lambda-\lambda z)]\frac{1}{\lambda(1-z)},\qquad i=1,2 \tag{32}$$

$$II)Q(z)=Q(0,z)[1-V^*(\lambda-\lambda z)]\frac{1}{\lambda(1-z)} \tag{33}$$

$$III)\quad P_2(0,z)=p_2P_2(0,z)B_2^*(\lambda-\lambda z)+q_1P_1(0,z)B_1^*(\lambda-\lambda z) \tag{34}$$

and

$$IV)Q(0,z)=\theta q_2P_2(0,z)B_2^*(\lambda-\lambda z) \tag{35}$$

$$V)\quad P_1(0,z)=\frac{\lambda R_0(z-1)}{z-\{p_1zB_1^*+[(1-\theta)+\theta V^*]\frac{q_1q_2B_1^*B_2^*}{1-p_2B_2^*}\}} \tag{36}$$

Proof. *I)By integration from $x=0$ to $x=+\infty$ in (28) we have I).*

II)By integration from $x=0$ to $x=+\infty$ in (29) we have II).

III)By multiplying (25) in z^n and summation from n=0 to ∞ the result is obtained .

IV)By using the same technique on (26), we have IV) .

V)First we multiply (24) in z^n and summation from n=1 to ∞, then using (22)and (23) we obtain

$$\begin{aligned}zP_1(0,z)=\lambda R_0(z-1)+p_1zP_1(0,z)B_1^*(\lambda-\lambda z)\\+(1-\theta)q_2P_2(0,z)B_2^*(\lambda-\lambda z)+Q(0,z)V^*(\lambda-\lambda z)\end{aligned} \tag{37}$$

Now from (34) and (35) we obtain V). □

In the rest of this section for simplification we omit $(\lambda-\lambda z)$.

Corollary 3.1.

By using (36) in (32)and (33) we have

$$P_1(z)=\frac{R_0(1-B_1^*)}{\{p_1zB_1^*+[(1-\theta)+\theta V^*]\frac{q_1q_2B_1^*B_2^*}{1-p_2B_2^*}\}-z} \tag{38}$$

$$P_2(z) = \frac{R_0 q_1 B_1^*(1-B_2^*)}{(1-p_2B_2^*)\{p_1 z B_1^* + [(1-\theta)+\theta V^*]\frac{q_1 q_2 B_1^* B_2^*}{1-p_2B_2^*}\} - z} \tag{39}$$

$$Q(z) = \frac{R_0 q_1 q_2 B_1^* B_2^*(1-V^*)}{(1-p_2B_2^*)\{p_1 z B_1^* + [(1-\theta)+\theta V^*]\frac{q_1 q_2 B_1^* B_2^*}{1-p_2B_2^*}\} - z} \tag{40}$$

Remark 3.1. The unknown constant R_0 can be determined by using normalizing condition (27) which is

$$R_0 + P_1(1) + P_2(1) + Q(1) = 1 \tag{41}$$

from (38),(39) and (40) by using L'Hopital rule we have

$$P_1(1) = R_0 \frac{\frac{\lambda E(B_1)}{q_1}}{1-\lambda[\frac{E(B_1)}{q_1}+\frac{E(B_2)}{q_2}+\theta E(V)]}$$
$$P_2(1) = R_0 \frac{\frac{\lambda E(B_2)}{q_2}}{1-\lambda[\frac{E(B_1)}{q_1}+\frac{E(B_2)}{q_2}+\theta E(V)]}$$
$$Q(1) = R_0 \frac{\lambda\theta E(V)}{1-\lambda[\frac{E(B_1)}{q_1}+\frac{E(B_2)}{q_2}+\theta E(V)]}$$

hence by substituting the above values in (36) and simplifying we have $R_0 = 1-\rho$ where

$$\rho = \lambda[\frac{E(B_1)}{q_1} + \frac{E(B_2)}{q_2} + \theta E(V)] \tag{42}$$

R_0 is the steady-state probability that the server is idle but available in the system, hence $\rho < 1$ can be the stability condition under which the steady state solution exists.

Now the PGF of the queue size distribution at a random epoch is

$$\begin{aligned} P(z) &= P_1(z) + P_2(z) + Q(z) \\ &= R_0\{\frac{(1-p_2B_2^*)(1-B_1^*) + q_1B_1^*(1-B_2^*) + \theta q_1 q_2 B_1^* B_2^*(1-V^*)}{(1-p_2B_2^*)\{p_1 z B_1^* + [(1-\theta)+\theta V^*]\frac{q_1 q_2 B_1^* B_2^*}{1-p_2B_2^*} - z\}} \\ &= R_0 \frac{1-p_1B_1^*-p_2B_2^*+[p_2-q_1+\theta q_1 q_2(1-V^*)]B_1^*B_2^*}{(1-p_2B_2^*)\{p_1 z B_1^* + [(1-\theta)+\theta V^*]\frac{q_1 q_2 B_1^* B_2^*}{1-p_2B_2^*} - z\}} \end{aligned} \tag{43}$$

The PGF of the queue size distribution at departure epoch is

$$\begin{aligned} P_Q(z) &= R_0 + zP(z) \\ &= R_0 \frac{(1-z)q_1q_2[(1-\theta)+\theta V^*]B_1^*B_2^*}{(1-p_2B_2^*)\{p_1zB_1^* + [(1-\theta)+\theta V^*]\frac{q_1q_2B_1^*B_2^*}{1-p_2B_2^*} - z\}} \\ &= R_0 \frac{(z-1)q_1q_2[(1-\theta)+\theta V^*]B_1^*B_2^*}{z(1-p_1B_1^*-p_2B_2^*+p_1p_2B_1B_2^*) - q_1q_2[(1-\theta)+\theta V^*]B_1^*B_2^*} \end{aligned} \tag{44}$$

4. Mean queue size and other measures of system

Let L_Q be the mean number of customers in the system (i.e mean queue size), then we have

$$L_Q = \frac{dP_Q(z)}{dz} \mid_{z=1} \tag{45}$$

Proposition 4.1. *From (45) and using (44) we have*

$$\begin{aligned} L_Q = \lambda E(B) + \frac{\lambda^2 E(B^2)}{2(1-\rho)} \\ + \frac{\lambda[\frac{p_1}{q_1}E(B_1)+\frac{p_2}{q_2}E(B_2)]}{1-\rho} + \frac{\lambda^2[\frac{p_1}{q_1}E(B_1^2)+\frac{p_2}{q_2}E(B_2^2)]}{2(1-\rho)} \\ + \lambda^2\frac{p_1p_2}{q_1q_2}E(B_1)E(B_2) \end{aligned} \tag{46}$$

where $E(B)$ and $E(B^2)$ are in (3) and (4).

Proof. *$P_Q(z)$ has the form $R_0\frac{f(z)}{g(z)}$, where*

$$f(z) = (z-1)q_1q_2[(1-\theta)+\theta V^*]B_1^*B_2^*$$

and

$$g(z) = z(1-p_1B_1^*-p_2B_2^*+p_1p_2B_1B_2^*) - q_1q_2[(1-\theta)+\theta V^*]B_1^*B_2^*$$

Since $\lim_{z\to 1} f(z) = \lim_{z\to 1} g(z) = 0$, then by using L'Hopital's rule , we have

$$L_Q = R_0\frac{f''(1)g'(1) - g''(1)f'(1)}{2[g'(1)]^2} \tag{47}$$

where $R_0 = 1-\rho$ and ρ is obtained from (42).By calculating $f^{'}(1)$,$f^{''}(1)$,$g^{'}(1)$, $g^{''}(1)$ and substituting in (47) we have (46). □

Now for computing the mean response time of a test customer in this model , we use the approach of Kleinrock [7].Let $W_Q^*(s)$ be the LST of DF of waiting time of a tagged customer in this model.Then we have

$$W_Q^*(\lambda-\lambda z)B^*(\lambda-\lambda z) = P_Q(z) \tag{48}$$

where B^* is defined in (2).
If W_R denotes the time interval from arrival time to the time when a tagged customer leaves the system after the completion of service, i.e waiting time plus service time, then

$$W_R^*(s) = W_Q^*(s)B^*(s) \tag{49}$$

and mean response time of a tagged customer is

$$E(W_R) = -\frac{dW_R^*(s)}{ds} \mid_{s=0} \tag{50}$$

By substituting from (48) in (49) we have

$$W_R^*(s) = P_Q(1 - \frac{s}{\lambda}) \tag{51}$$

By using (44) and from (50) we have

$$E(W_R) = \frac{1}{\lambda}L_Q \tag{52}$$

Also the average system size is $L = L_Q + \rho$ where ρ is in (42).

5. Particular cases

i)If we set $p_i = 0$ and $q_i = 1$ for all $i = 1, 2, ..., k$, i.e there isn't feedback in each phase, then we obtain the results of [12]. Also with $k = 2$ the results of [8] are obtained.

ii)If we set $\theta = 0$, i.e there isn't vacation in system, then from (44) we have

$$P_Q(z) = R_0\frac{(z-1)q_1q_2B_1^*B_2^*}{z(1 - p_1B_1^* - p_2B_2^* + p_1p_2B_1B_2^*) - q_1q_2B_1^*B_2^*} \tag{53}$$

where $R_0 = 1 - \rho$ and $\rho = \lambda[\frac{E(B_1)}{q_1} + \frac{E(B_2)}{q_2}]$.Also L_Q is obtained from (46) with $E(B) = E(B_1) + E(B_2)$ and $E(B^2) = E(B_1^2) + E(B_2^2) + 2E(B_1)E(B_2)$
iii)Finally with $\theta = 0$,$p_1 = p_2 = 0$, $q_1 = q_2 = 1$ and $B_2^* = 0$ then we have

$$P_Q(z) = R_0\frac{(z-1)B_1^*}{z - B_1^*} \tag{54}$$

where $R_0 = 1 - \rho$ and $\rho = \lambda E(B_1)$ and

$$L_Q = \lambda E(B_1) + \frac{\lambda^2 E(B_1^2)}{2(1-\rho)} \tag{55}$$

which is the famous Pollaczek-Khinchine formula.

References

1. Badamchizadeh.A; Shahkar.G.H, A two phase queue system with Bernoulli feedback and Bernoulli schedule server vacation, Information and Management Sciences, Vol 19, No.2, 2008
2. Doo Choi;Tae-Sung Kim, Analysis of a Two- phase Queueing system with Vacation and Bernoulli feedback, Stocastic analysis and applications,Vol 21, No.5, pp.1009-1019,2003
3. G.Choudhury ;M.Paul, A two phase queueing system with Bernoulli feedback, Information and Managment Science ,Vol16,No1,pp35-52,2005.
4. B.T.Doshi, Queueing sysytem with vacation - A survey, Queueing System, vol 1 (1986)29-66
5. J.Keilson ; L.D.Servi, Oscillating random walk models for $G1/G/1$ vacation systems with Bernoulli schedules, Journal of Applied Probability 23 (1986), 790-802.
6. J.Keilson ; L.D.Servi, Dynamic of the M/G/1 vacation model, Operation Research ,Vol 35,No4,July-Agust 1987.
7. Kleinrock,L., Queueing System;Wiley:New York,1975;Vol.1.
8. K.C.Madan; G.Choudhury,A single server queue with two phases of heterogenous service under Bernoulli schedule and a general vcation time,Information and Managment Science, vol 16,Number 2,pp1-16,2005
9. R.Ramaswam ; L.D.Servi, The busy period of the $M/G/1$ vacation model with a Bernoulli schedule. Stochastic Modeles 4 (1988)507-521
10. Salehirad.M; Badamchizadeh.A, On the multi-phase $M/G/1$ queueing system with random feedback, CEJOR, Vol 17, pp.131-139,2009.
11. L.D.Servi, Average delay approxmation of $M/G/1$ cyclic service queues with Bernoulli schedules,IEEE Selected Areas of Communications 4(6)(1986)813-820(Correction in 5(3)(1987)547).
12. G.H.Shahkar; A.Badamchi Zadeh, On $M/(G_1, G_2, ..., G_k)/V/1/(BS)$. Far East J.Theo.Stat 20(2)2006,151-162.
13. H.Takagi, Queueing Analysis :A Foundation of Performance Evalution. vol 1, North Holland, Amesterdam, 1991.

Asymptotic Behaviors of the Lorenz Curve for Left Truncated and Dependent Data

M. Bolbolian Ghalibaf, V. Fakoor and H.A. Azarnoosh

Department of Statistics, Ferdowsi University of Mashhad

In this paper, we consider a nonparametric estimator of the Lorenz curve under truncated dependent model. We show that this estimator is uniformly strongly consistent for the associated Lorenz curve. Also, a strong Gaussian approximation for the associated Lorenz process are established under appropriate assumptions. A law of the iterated logarithm for the Lorenz process is also derived.

Keywords: Law of the iterated logarithm, Lorenz curve, Quantile function, Strong consistency, Strong Gaussian approximation, Strong mixing, Truncated data

1. Introduction and Preliminaries

Pietra (1915) and Gastwirth (1971) independently introduced the *Lorenz curve* corresponding to a non-negative random variable (rv) X with a distribution function (df) F, quantile function $Q(p)$ and finite mean $EX = \mu$ as:

$$L_F(t) := \frac{1}{\mu}\int_0^t Q(s)ds, \quad 0 \le t \le 1.$$

In econometrics, with X representing income, $L(t)$ gives the fraction of total income that the holders of the lowest t^{th} fraction of income possesses. Most of the measures of income inequality are derived from the Lorenz curve. An important example is the Gini index associated with F defined by

$$G_F := \frac{\int_0^1 [u - L_F(u)]du}{\int_0^1 udu} = 1 - 2(CL)_F,$$

where $(CL)_F = \int_0^1 L_F(u)du$ is the *cumulative Lorenz curve* corresponding to F. This is a ratio of the area between the Lorenz curve and the 45° line to the area under the 45° line. The numerator is usually called the *area of concentration*. Kendall and Stuart (1963) showed that this is equivalent to a ratio of a measure of dispersion to the mean. In general, these notions are useful for measuring concentration and inequality in distributions of resources, and in size distributions. For

a list of applications in different areas, we refer the readers to Csörgő and Zitikis (1996a).

To estimate the Lorenz curve, one can use the *Lorenz statistic* $L_n(y)$ defined by

$$L_n(y) := \frac{1}{\mu_n} \int_0^y Q_n(u)du, \quad 0 \leq y \leq 1,$$

where μ_n is the sample mean and $Q_n(y)$ is the empirical quantile function constructed from i.i.d. sample taken from F.

Goldie (1977) proved the uniform consistency of L_n to L_F and derived the weak convergence of the *Lorenz process* $l_n(t) := \sqrt{n}[L_n(t) - L(t)], 0 \leq t \leq 1$ to a Gaussian process under suitable conditions. Csörgő et al. (1986) gave a unified treatment of strong and weak approximations of the Lorenz and other related processes. In particular, they established a strong invariance principle for the Lorenz process, by which Rao and Zhao (1995) derived one of their two versions of the law of the iterated logarithm (LIL) for the Lorenz process. Different versions of the LIL under weaker assumptions are also obtained by Csörgő and Zitikis (1996a, 1997). In Csörgő and Zitikis (1996b), confidence bands for the Lorenz curve that are based on weighted approximations of the Lorenz process are constructed. Csörgő et al. (1987), obtained weak approximations for Lorenz curves under random right censorship. Strong Gaussian approximations for the Lorenz process when data are subject to random right censorship and left truncation are established by Tse (2006), he is also derived a functional LIL for the Lorenz process.

However, in most economic situations, the basic sequence of observations may not be independent. It is more realistic to assume some form of dependence among the data are observed. Csörgő and Yu (1999), obtained weak approximations for Lorenz curves and its inverse under the assumption of mixing dependence. Glivenko-Cantelli-type asymptotic behavior of the empirical generalized Lorenz curves based on random variables forming a stationary ergodic sequence with deterministic noise were considered by Davydov and Zitikis (2002). Davydov and Zitikis (2003) established a large sample asymptotic theory for the empirical generalized Lorenz curves when observations are stationary and either short-range or long-range dependent. Strong laws for the generalized absolute Lorenz curves when data are stationary and ergodic sequences established by Helmers and Zitikis (2005). Based on the generalized Lorenz curves Davydov et al. (2007) proposed a statistical index for measuring the fluctuations of a stochastic process. They developed some of the asymptotic theory of the statistical index in the case where the stochastic process is a Gaussian process with stationary increments and a nicely behaved correlation function. The uniform strong convergence rate of the Lorenz curve estimator under strong mixing hypothesis is obtained by Fakoor et

al. (2009). They also established a strong Gaussian approximation for the Lorenz process, by which they derived a functional LIL for the Lorenz process, under the assumption of strong mixing. The counterpart of these results for the censored dependent model was established by Bolbolian et al. (2009a).

The purpose of this paper is to provide some asymptotic results for Lorenz process $l_n(t)$, for the case in which data are assumed to be strong mixing subject to random left truncation.

Consider a sequence of rv's $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_N}$ with common unknown absolutely continuous df F and finite mean μ. These rv's are regarded as the lifetimes of the items under study which may not be mutually independent. Among the different forms in which incomplete data appear, right censoring and left truncation are two common ones. Left truncation may occur if the time origin of the lifetime precedes the time origin of the study. Only subjects that fail after the start of the study are being observed, otherwise they are left truncated. This means that some subjects are sampled, while others are neglected. This model arises in various fields, e.g., astronomy, economy and medical studies (see, e.g., Woodroofe, 1985). Let $\mathbf{T_1}, \mathbf{T_2}, \ldots, \mathbf{T_N}$ be a sequence of independent and identically distributed (iid) random variables with continuous df G, they are also assumed to be independent of the rv's $\mathbf{X_i}$'s. In the left truncation model, $(\mathbf{X_i}, \mathbf{T_i})$ is observed only when $\mathbf{X_i} \geq \mathbf{T_i}$. Let $(X_1, T_1), \ldots, (X_n, T_n)$ be a sample which one observes (*i.e.*, $X_i \geq T_i$), and $\gamma := \mathbf{P}(\mathbf{T_1} \leq \mathbf{X_1}) > 0$, where $\mathbf{P}$ is the absolute probability (related to the N-sample). Note that n itself is a rv and that γ can be estimated by n/N (although this estimator cannot be calculated since N is unknown). Assume, without loss of generality, that $\mathbf{X_i}$ and $\mathbf{T_i}$ are nonnegative random variables, $i = 1, \ldots, N$. For any df L denotes the left and right endpoints of its support by $a_L = \inf\{x : L(x) > 0\}$ and $b_L = \sup\{x : L(x) < 1\}$, respectively. Then under the current model, as discussed by Woodroofe (1985), we assume that $a_G \leq a_F$ and $b_G \leq b_F$. Define

$$C(x) = \mathbf{P}(\mathbf{T_1} \leq x \leq \mathbf{X_1} | \mathbf{T_1} \leq \mathbf{X_1}) = \mathbb{P}(T_1 \leq x \leq X_1) = \gamma^{-1} G(x)(1 - F(x)), \quad (1)$$

where $\mathbb{P}(.) = \mathbf{P}(.|n)$ is the conditional probability (related to the n-sample) and consider its empirical estimate

$$C_n(x) = n^{-1} \sum_{i=1}^{n} I(T_i \leq x \leq X_i), \quad (2)$$

where $I(.)$ is the indicator function. Then the product-limit (PL) estimator $\widehat{F}_n$ of F is given by

$$\widehat{F}_n(x) = 1 - \prod_{X_i \leq x} \left(1 - \frac{1}{nC_n(X_i)}\right). \quad (3)$$

The cumulative hazard function $\Lambda(x)$ is defined by

$$\Lambda(x) = \int_0^x \frac{dF(u)}{1-F(u)}. \tag{4}$$

Let

$$F^*(x) = \mathbf{P}(\mathbf{X_1} \leq x | \mathbf{T_1} \leq \mathbf{X_1}) = \mathbb{P}(X_1 \leq x) = \gamma^{-1} \int_0^x G(u) dF(u), \tag{5}$$

be the df of the observed lifetimes. Its empirical estimator is given by

$$F_n^*(x) = n^{-1} \sum_{i=1}^n I(X_i \leq x).$$

On the other hand, the df of the observed T_i's is given by

$$G^*(x) = \mathbf{P}(\mathbf{T_1} \leq x | \mathbf{T_1} \leq \mathbf{X_1}) = \mathbb{P}(T_1 \leq x) = \gamma^{-1} \int_0^\infty G(x \wedge u) dF(u),$$

and is estimated by

$$G_n^*(x) = n^{-1} \sum_{i=1}^n I(T_i \leq x).$$

It then follows from (1) and (2) that

$$C(x) = G^*(x) - F^*(x), \quad C_n(x) = G_n^*(x) - F_n^*(x-). \tag{6}$$

Finally (1), (4) and (5) give

$$\Lambda(x) = \int_0^x \frac{dF^*(u)}{C(u)}.$$

Hence, a natural estimator of Λ is given by

$$\widehat{\Lambda}_n(x) = \int_0^x \frac{dF_n^*(u)}{C_n(u)} = \sum_{i=1}^n \frac{I(X_i \leq x)}{nC_n(X_i)},$$

which is the usual so-called Nelson-Aalen estimator of Λ. Moreover, $\widehat{\Lambda}_n$ is the cumulative hazard function of the PL estimator $\widehat{F}_n$ defined in (3).

The quantile function Q and its empirical counterpart Q_n are defined by

$$Q(p) = \inf\{x \in R; F(x) \geq p\} \quad \textit{and} \quad Q_n(p) = \inf\{x \in R; \widehat{F}_n(x) \geq p\} \tag{7}$$

Suppose that $0 < p_0 \leq p_1 < 1$.

We defined the Lorenz curve corresponding to rv X as:

$$L_F(p) := \frac{1}{\mu} \int_{p_0}^p Q(s) ds, \quad p_0 \leq p \leq p_1,$$

where $\mu = \int_{p_0}^{p_1} Q(s) ds$.

Therefore the natural estimator for the Lorenz curve $L_F(t)$ is

$$L_n(p) := \frac{1}{\mu_n} \int_{p_0}^{p} Q_n(s)ds, \quad p_0 \leq p \leq p_1,$$

where $\mu_n = \int_{p_0}^{p_1} Q_n(s)ds$.

The main aims of this paper are to derive strong uniform consistency of the Lorenz statistic and strong Gaussian approximation for Lorenz process, for the case in which data are assumed to be dependent subject to random left truncation. As a result of our strong Gaussian approximation, we obtain a functional LIL for the Lorenz process.

In this paper we consider the strong mixing dependence, which amounts to a form of asymptotic independence between the past and the future as shown by its definition.

Definition 1.1. Let $\{X_i, i \geq 1\}$ denote a sequence of random variables. Given a positive integer m, set

$$\alpha(m) = \sup_{k \geq 1} \{|P(A \cap B) - P(A)P(B)| \ ; \ A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+m}^{\infty}\}, \qquad (8)$$

where $\mathcal{F}_i^k$ denote the σ-field of events generated by $\{X_j; i \leq j \leq k\}$. The sequence is said to be strong mixing (α-mixing) if the mixing coefficient $\alpha(m) \to 0$ as $m \to \infty$.

Among various mixing conditions used in the literature, strong mixing is reasonably weak and has many practical applications (see, e.g. Doukhan (1994) or Cai (1998, 2001) for more details). In particular, Masry and Tjostheim (1995) proved that, both ARCH processes and nonlinear additive AR models with exogenous variables, which are particularly popular in finance and econometrics, are stationary and strong mixing.

Now we introduce our main assumption that is used to state our results gathered below for easy reference.

A. $\{X_i\}_{i \geq 1}$ is a sequence of stationary strong mixing rv's with mixing coefficient $\alpha(n) = O(e^{-(\log n)^{1+\nu}})$ for some $\nu > 0$.

In the next Section, we present our main results.

2. Asymptotic Behaviors of Lorenz Curve

2.1. *Strong Uniform Consistency*

Theorem 2.1 below proves the uniform strong consistency with rate of the estimator L_n.

Theorem 2.1. *Let $0 < p_0 \leq p_1 < 1$. Under Assumption **A**, assuming that $F' = f$ is bounded away from zero on $[Q(p_0) - \delta, Q(p_1) + \delta)$, for some $\delta > 0$. Then*

$$\sup_{p_0 \leq p \leq p_1} |L_n(p) - L_F(p)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s. \tag{9}$$

Proof. An elementary computation shows that,

$$L_n(p) - L_F(p) = \frac{1}{\mu_n}\int_{p_0}^{p}[Q_n(s) - Q(s)]ds - \frac{\mu_n - \mu}{\mu_n}L_F(p). \tag{10}$$

It is easy to see that,

$$\mu_n - \mu = \int_{p_0}^{p_1}[Q_n(s) - Q(s)]ds. \tag{11}$$

Now, by using (10), (11) and Lemma 3 of Lemdani et al. (2005), we obtain the results. □

2.2. *Strong Gaussian Approximation*

We first introduce the following Gaussian process, which plays an important role to present our strong approximation.

Let $g_j(s) = I(X_j \leq s) - F^*(s), \quad j \geq 0,$

$$\Gamma(s, s') = Cov(g_1(s), g_1(s')) + \sum_{j=2}^{\infty}[Cov(g_1(s), g_j(s')) + Cov(g_1(s'), g_j(s))]. \tag{12}$$

Define, for $0 \leq t \leq b$, two parameter mean zero Gaussian process

$$B(t, n) := \frac{K(t, n)/\sqrt{n}}{C(t)} + \int_0^t \frac{K(u, n)/\sqrt{n}}{C^2(u)}dC(u), \tag{13}$$

where $\{K(s, t), 0 \leq s, t \leq b\}$ is a Kiefer process in Theorem 3 of Dhompongsa (1984) with covariance function

$$\Gamma^*(t, t', s, s') = \min(t, t')\Gamma(s, s'),$$

and $\Gamma(s, s')$ given by (12).

We now restate below a strong approximation by Bolbolian et al. (2009b) for the normed PL-quantile process $\rho_n(u) := \sqrt{n}f(Q(u))[Q(u) - Q_n(u)]$ by a

two parameter Gaussian process at the rate $O((\log n)^{-\lambda})$, for some $\lambda > 0$. The statements are conditional on the observed sample size n.

Theorem 2.2. *(Bolbolian et al., 2009b) Let $0 < p_0 \leq p_1 < 1$. Under Assumption **A**, assume that F is Lipschtiz continuous and that F is twice continuously differentiable on $[Q(p_0) - \delta, Q(p_1) + \delta]$, for some $\delta > 0$, such that f is bounded away from zero, then there exists a two parameter mean zero Gaussian process $B(x, u)$ for $x, u \geq 0$, such that,*

$$\sup_{p_0 \leq p \leq p_1} |\rho_n(p) - (1-p)B(Q(p), n)| = O((\log n)^{-\lambda}) \quad a.s.,$$

for some $\lambda > 0$. $\square$

We will give strong Gaussian approximation of the Lorenz process over restricted interval $[p_0, p_1]$ for fixed $0 < p_0 \leq p_1 < 1$.

In the full model, Langberg et al. (1980) define the *total time on test transform* curve corresponding to a continuous distribution F on $[0, \infty)$, $H_F^{-1}(p)$, for $p \in [0, 1]$ as

$$H_F^{-1}(p) = \int_0^p (1-y)dQ(y) = (1-p)Q(p) + \int_0^p Q(y)dy, \quad Q(0) = 0.$$

Obviously, $H_F^{-1}(p) \leq H_F^{-1}(1) := \lim_{p \uparrow 1} H_F^{-1}(p) = \mu$. For the our model, we modify the definition of $H_F^{-1}(p)$ as

$$H_F^{-1}(p) = (p_1 - p)Q(p) + \int_{p_0}^p Q(y)dy, \quad p \in [p_0, p_1]. \tag{14}$$

As $p_0 \downarrow 0$ and $p_1 \uparrow 1, H_F^{-1}(p_1) \to \int_0^1 Q(y)dy = \mu$. We can regard $H_F^{-1}(p_1)$ as a surrogate for the finite mean μ. A natural estimator for $H_F^{-1}(p)$ is

$$H_n^{-1}(p) = (p_1 - p)Q_n(p) + \int_{p_0}^p Q_n(y)dy, \quad p \in [p_0, p_1].$$

In the next theorem, we construct a two parameter mean zero Gaussian process that strongly uniformly approximate the empirical process $l_n(p)$.

Theorem 2.3. *Let $0 < p_0 \leq p_1 < 1$. Under Assumption **A**, assume that F is Lipschtiz continuous and that F is twice continuously differentiable on $[Q(p_0) - \delta, Q(p_1) + \delta]$, for some $\delta > 0$, such that f is bounded away from zero. Then there exists a two parameter mean zero Gaussian process $B(t, u)$ for $t, u \geq 0$, such that,*

almost surely,

$$\sup_{p_0\le p\le p_1}\Bigg| l_n(p) - \frac{1}{H_F^{-1}(p_1)}\left(\int_{p_0}^{p}\frac{(p_1-y)B(Q(y),n)}{f(Q(y))}dy - L_F(p)\int_{p_0}^{p_1}\frac{(p_1-y)B(Q(y),n)}{f(Q(y))}dy\right)\Bigg| = O((\log n)^{-\lambda}), \quad (15)$$

for some $\lambda > 0$.

Proof. See the Appendix. □

2.3. *Functional LIL*

The next theorem gives a functional LIL for the Lorenz process. We work on the probability space of Theorem 2.3. Let $D[a,b]$ be the space of functions on $[a,b]$ that are right continuous and have left limits and B is the unit ball in the reproduce kernel Hilbert space $H(\Gamma^*)$.

Theorem 2.4. *Suppose that conditions of Theorem 2.3 are satisfied. On a rich enough probability space,* $l_n(.)/\sqrt{2\log\log n}$ *is almost surly relatively compact in* $D[p_0,p_1]$ *with respect to the supremum norm and its set of limit points is*

$$\mathrm{G} = \Bigg\{ g_h : g_h(u) = \frac{1}{H_F^{-1}(p_1)}\left(\int_{p_0}^{u}\frac{h(y)}{f(Q(y))}dy - L_F(u)\int_{p_0}^{p_1}\frac{h(y)}{f(Q(y))}dy\right), \quad p_0\le u\le p_1,\ h\in\mathcal{H}\Bigg\},$$

where

$$\mathcal{H} = \left\{ h : [p_0,p_1]\to \boldsymbol{R},\quad h(u) = \frac{g(u)}{C(u)} + \int_0^u \frac{g(x)}{C^2(x)}dC(x) : g\in \mathrm{B}\right\}.$$

Proof. Theorem 2.4 follows at once from (15) and Theorem A in Berkes and Philipp (1977). □

3. Appendix

In establishing Theorem 2.3, we were aided by some ideas found in Tse (2006), but first we start with the following lemmas which is necessary for achieving the establishment of the our results.

Lemma 3.1. *Suppose the conditions of Theorem 2.2 are satisfied. We have,*

$$\lim_{n\to\infty} \sup_{p_0\le p\le p_1} |H_n^{-1}(p) - H_F^{-1}(p)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s.$$

Proof. By Lemma 3 of Lemdani et al. (2005), we have,

$$\begin{aligned}\sup_{p_0\le p\le p_1} |H_n^{-1}(p) - H_F^{-1}(p)| \le & \sup_{p_0\le p\le p_1} [(p_1-p)|Q_n(p) - Q(p)|] \\ & + \sup_{p_0\le p\le p_1} \int_{p_0}^{p} |Q_n(y) - Q(y)|dy \\ & = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s. \qquad \square\end{aligned}$$

Next, define the normed total time on test empirical process $t_n(p)$ by

$$t_n(p) = \sqrt{n}[H_n^{-1}(p) - H_F^{-1}(p)], \quad p \in [p_0, p_1].$$

Lemma 3.2 characterize the asymptotic limit of $t_n(p)$.

Lemma 3.2. *Suppose the conditions of Theorem 2.2 are satisfied. Then there exists a two parameter mean zero Gaussian process $B(t,u)$ for $t,u \ge 0$, such that,*

$$\begin{aligned}\sup_{p_0\le p\le p_1} \left| t_n(p) - \left(\int_{p_0}^{p} \frac{(p_1-y)B(Q(y),n)}{f(Q(y))}dy + \frac{(p_1-p)^2 B(Q(p),n)}{f(Q(p))} \right) \right| & \\ = O((\log n)^{-\lambda}) \quad a.s.\end{aligned}$$

Proof. Proof of this lemma can be done using similar augment of Lemma 3.2 in Tse (2006), we therefore omit the proof. □

Next, we define the scaled total time on test transform, its statistic and associated empirical process corresponding to F.

$$W_F(p) := \frac{H_F^{-1}(p)}{H_F^{-1}(p_1)}, \quad W_n(p) := \frac{H_n^{-1}(p)}{H_n^{-1}(p_1)} \tag{16}$$

and

$$w_n(p) := \sqrt{n}[W_n(p) - W_F(p)]$$

for $p \in [p_0, p_1]$.

The following lemmas give the strong uniform consistency of $W_n(p)$ and strong Gaussian approximation of the scaled total time on test empirical process respectively.

Lemma 3.3. *Suppose that conditions of Theorem 2.2 are satisfied. We have,*

$$\sup_{p_0\le p\le p_1}|W_n(p)-W_F(p)|=O\left(\sqrt{\frac{\log\log n}{n}}\right)\quad a.s.$$

Proof. By triangular inequality and Lemma 3.1, the left hand side is bounded by

$$\begin{aligned}&\sup_{p_0\le p\le 1}\left|\frac{H_n^{-1}(p)}{H_n^{-1}(p_1)}-\frac{H_n^{-1}(p)}{H_F^{-1}(p_1)}\right|+\sup_{p_0\le p\le p_1}\left|\frac{H_n^{-1}(p)}{H_F^{-1}(p_1)}-\frac{H_F^{-1}(p)}{H_F^{-1}(p_1)}\right|\\ &\le\sup_{p_0\le p\le p_1}\left|H_n^{-1}(p)\frac{H_F^{-1}(p_1)-H_n^{-1}(p_1)}{H_n^{-1}(p_1)H_F^{-1}(p_1)}\right|+\sup_{p_0\le p\le p_1}\left|\frac{1}{H_F^{-1}(p_1)}[H_F^{-1}(p)-H_n^{-1}(p)]\right|\\ &=O\left(\sqrt{\frac{\log\log n}{n}}\right)\quad a.s.\qquad\square\end{aligned}$$

Lemma 3.4. *Suppose that conditions of Theorem 2.2 are satisfied. Then there exists a two parameter mean zero Gaussian process $B(t,u)$ for $t,u\ge 0$, such that,*

$$\sup_{p_0\le p\le p_1}|w_n(p)-\frac{1}{H_F^{-1}(p_1)}\left(\int_{p_0}^{p}\frac{(p_1-y)B(Q(y),n)}{f(Q(y))}dy+\frac{(p_1-p)^2B(Q(p),n)}{f(Q(p))}\right)$$

$$+\frac{H_F^{-1}(p)}{(H_F^{-1}(p_1))^2}\int_{p_0}^{p_1}\frac{(p_1-y)B(Q(y),n)}{f(Q(y))}dy|=O((\log n)^{-\lambda})\quad a.s.$$

for some $\lambda>0$.

Proof. Proof can be done along the lines of Lemma 3.5 of Tse (2006), we therefore omit the proof. □

Proof of Theorem 2.3. By Definition of the Lorenz curve corresponding to F in the our model and by using (14) and (16) we have

$$W_F(y)=\frac{(p_1-y)Q(y)}{\int_{p_0}^{p_1}Q(u)du}+L_F(y).\tag{17}$$

We have also

$$W_n(y)=\frac{(p_1-y)Q_n(y)}{\int_{p_0}^{p_1}Q_n(p)dp}+L_n(y),\quad y\in[p_0,p_1].\tag{18}$$

Substituting (17) and (18) in Lemma 3.4, we obtain the result. □

References

1. Berkes, I. and Philipp, W. (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrscheinlichkeitstheorie verw.* Gebiete **41**, 115-137.
2. Bolbolian Ghalibaf, M., Fakoor, V. and Azarnoosh, H.A. (2009a). Asymptotic behaviors of the Lorenz curve for censored data under strong mixing. *Communications in Statistics - Theory and Methods.* In press.
3. Bolbolian Ghalibaf, M., Fakoor, V. and Azarnoosh, H.A. (2009b). Strong Gaussian approximations of product-limit and quantile processes for truncated data under strong mixing. *Statistics and Probability Letters.* In press.
4. Cai, Z. (1998). Kernel density and hazard rate estimation for censored dependent data. *Journal of Multivariate Analysis.* **67**, 23-34.
5. Cai, Z. (2001). Estimating a distribution function for censored times series data. *Journal of Multivariate Analysis.* **78**, 299-318.
6. Csörgő, M., Csörgő, S. and Horváth, L. (1986). *An asymptotic theory for empirical reliability and concentration processes.* Lecture Notes in Statistics, Vol. **33**, Springer Berlin Heidelberg NewYork.
7. Csörgő, M., Csörgő, S. and Horváth, L. (1987). Estimation of total time on test transforms and Lorenz curves under random censorship. *Statistics.* **18**, 77-97.
8. Csörgő, M. and Yu, H. (1999). Weak approximations for empirical Lorenz curves and their Goldie inverses of stationary observations. *Adv. Appl. Prob.* **31**, 698-719.
9. Csörgő, M. and Zitikis, R. (1996a). Strassens LIL for the Lorenz curve. *Journal of Multivariate Analysis.* **59**, 1-12.
10. Csörgő, M. and Zitikis, R. (1996b). *Confidence bands for the Lorenz curve and Goldie curves.* In A volume in honor of Samuel Kotz. NewYork: Wiley.
11. Csörgő, M. and Zitikis, R. (1997). On the rate of strong consistency of Lorenz curves. *Statistics and Probability Letters.* **34**, 113-121.
12. Davydov, Y. and Zitikis, R. (2002). Convergence of generalized Lorenz curves based on stationary ergodic random sequences with deterministic noise. *Statistics and Probability Letters.* **59**, 329-340.
13. Davydov, Y. and Zitikis, R. (2003). Generalized Lorenz curves and convexifications of stochastic processes. *J. Appl. Probab.* Vol. 40, N. **4**, 906-925.
14. Davydov, Y., Khoshnevisan, D., Shic, Z. and Zitikis, R. (2007). Convex rearrangements, generalized Lorenz curves, and correlated Gaussian data. *Journal of Statistical Planning and Inference.* **137**, 915-934.
15. Dhompongsa, S. (1984). A note on the almost sure approximation of the empirical process of weakly dependent random variables. *Yokohama Math. J.*

32, 113-121.

16. Doukhan, P. (1994). *Mixing: Properties and examples.* Lecture Notes in Statistics. **85**, Springer-Verlag, New York.
17. Fakoor, V. and Nakhaei Rad, N. (2009). Asymptotic behaviors of the Lorenz curve under strong mixing. Submitted.
18. Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica.* **39**, 1037-1039.
19. Goldie, C.M. (1977). Convergence theorems for empirical Lorenz curve and their inverses. *Advances in Applied Probability.* **9**, 765-791.
20. Helmers, R. and Zitikis, R. (2005). Strong laws for generalized absolute Lorenz curves when data are stationary and ergodic sequences. *Proc. Amer. Math. Soc.* **133**, 3703-3712.
21. Kendall, M.G. and Stuart, A. (1963). *The advanced theory of statistics I.* (2nd. ed.) Charles Griffen and Company, London.
22. Langberg, N.A. and Leon, R.V. and Proschan, F. (1980). Characterization of nonparametric classes of life distributions. *Annals of Probability.* **8**, 1163-1170.
23. Lemdani, M., Ould-Saïd, E. and Poulin, N. (2005). Strong representation of the quantile function for left truncated and dependent data. *Mathematical Methods of statistics.* Vol. 14, No. **3**, 332-345.
24. Masry, E. and Tjostheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econ. Theor.* **11**, 258-289.
25. Pietra, G. (1915). Delle relazioni fra indici di variabilitá, note I e II. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti.* **74,** 775-804.
26. Rao, C.R. and Zhao, L.C. (1995). Strassens law of the iterated logarithm for the Lorenz curves. *Journal of Multivariate Analysis.* **54**, 239-252.
27. Tse, S.M. (2006). Lorenz curve for truncated and censored data. *AISM.* **58**, 675-686.
28. Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163-177.

A Kolmogorov inequality for quadratic forms of NSD uniformly bounded random variables

N. Eghbal, M. Amini and A. Bozorgnia

Department of Statistics, Ferdowsi University of Mashhad

In this paper, we present a Kolmogorov probability inequality for quadratic forms and weighted quadratic forms of negative superadditive dependent (NSD) uniformly bounded random variables. Using these inequalities, we evaluate complete convergence of randomized quadratic forms under some suitable conditions. Moreover, various examples presented for which satisfy given conditions in paper.

Keywords: Kolmogorov inequality, Negative superadditive dependence, Quadratic forms, Complete convergence.

1. Introduction

Let $\{X_i; i \geq 1\}$ be a sequence of independent identically distributed random variables (r.v.'s). Consider the following quadratic forms (Q.F.'s)

$$Q_n = \sum_{1 \leq i < j \leq n} a_{ij} X_i X_j, \quad n \geq 2,$$

where $\{a_{ij}; 1 \leq i < j \leq n\}$ is an array of real numbers. In particular, if $a_{ij} = 1$ for all $i \neq j$, then we define

$$T_n = \sum_{1 \leq i < j \leq n} X_i X_j, \quad n \geq 2.$$

Many authors have been studied limiting behavior of the quadratic forms and weighted quadratic forms $\{T_n; n > 1\}$ and $\{Q_n; n \geq 1\}$ respectively. For instance, Cuzich et al. (1995), Zhang (1996), Whittle (1960, 1964), Varberg (1966) and Eghbal et al. (2010). Moreover, the study of U-statistics was initiated by Hoeffding (1948) and applications can be found in many references, e.g. Serfling (1980). Improvements, extensions and results related to Kolmogorov inequalities can be found among others, in Young et al. (1987), Turner et al. (1995) and Mavrikiou (2007). Exponential and Kolmogorov inequalities have been constructed for U-statistics based on Bernoulli kernels in Christofides (1991, 1994). Also Mavrikiou (2008) obtained a Kolmogorov inequality for weighted U-statistics. In this paper,

we derive a Kolmogorov inequality for quadratic forms, $T_n = \sum_{1\le i<j\le n} X_iX_j$ and weighted quadratic forms $W_n = \sum_{1\le i<j\le n} a_{ij}X_iX_j$, where $\{X_i; i \ge 1\}$ is a sequence of nonnegative NSD uniformly bounded random variables and $\{a_{ij}; 1 \le i < j \le n\}$ be an array of non-negative real numbers. Then, using these inequalities we evaluate complete convergence of randomized quadratic forms under some suitable conditions. Moreover, some well known multivariate distributions that possess the NSD property and satisfy in our conditions present in section 4.

2. Preliminaries

Here, We present some well known definitions and lemmas will be used in next section.

Definition 2.1. (Kemperman; 1977) A function $\phi : \mathcal{R}^m \to \mathcal{R}$ is called superadditive if $\phi(\mathbf{x} \vee \mathbf{y}) + \phi(\mathbf{x} \wedge \mathbf{y}) \ge \phi(\mathbf{x}) + \phi(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{R}^m$, where $\vee$ is for componentwise maximum and $\wedge$ is for componentwise minimum.

Lemma 2.1. *(Kemperman; 1977) If ϕ has continuous second partial derivatives, then the superadditivity of ϕ is equivalent to $\partial^2\phi/\partial x_i\partial x_j \ge 0$, $1 \le i \ne j \le m$.*

Definition 2.2. (Hu; 2000) A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ is said to be negatively superadditive dependent (*NSD*) if

$$E\phi(X_1, X_2, \ldots, X_m) \le E\phi(X_1^*, X_2^*, \ldots, X_m^*), \tag{1}$$

where $X_1^*, X_2^*, \ldots, X_m^*$ are independent with $X_i \overset{st}{=} X_i^*$ for each i and ϕ is a superadditive function such that the expectations in (1) exist. The following lemma that can be found in Hoeffding (1963) is based in our results.

Lemma 2.2. *Let X be a random variable such that $P(a \le X \le b) = 1, a < b$, and $E(X) = \mu$. Then,*

$$Ee^{t(X-\mu)} \le e^{\frac{1}{8}t^2(b-c)^2} \quad \text{for all } t > 0.$$

Noting that,without loss of generality, we suppose that $\{X_i; i \ge 1\}$ is a sequence of nonnegative random variables. Because, if assumption of nonnegativity random variables is removed, then by $|X| = X^+ + X^-$ and

$$\begin{aligned} |T_n| &\le \sum_{1\le i<j\le n} |X_iX_j| = \sum_{1\le i<j\le n} |X_i||X_j| \\ &= \sum_{1\le i<j\le n} X_i^+X_j^+ + \sum_{1\le i<j\le n} X_i^+X_j^- + \sum_{1\le i<j\le n} X_i^-X_j^+ + \sum_{1\le i<j\le n} X_i^-X_j^- \end{aligned}$$

all of our results are valid, where $X^+ = \max\{X, 0\}$ and $X^- = \max\{0, -X\}$.

3. Theoretical results

In this section, we derive a Kolmogorov inequality for quadratic forms of non-negative NSD uniformly bounded random variables. Also, we discuss the complete convergence of randomized quadratic forms under some suitable conditions.

Theorem 3.1. *Let $\{X_i; i \geq 1\}$ be a sequence of non-negative NSD random variables with $P(a \leq X_i \leq b) = 1$, $i = 1, 2, \cdots$. Then, for all $\varepsilon > 0$*

$$P\Big(T_n - E(T_n) \geq \varepsilon\Big) \leq \exp\left\{\frac{-4\varepsilon^2}{n(n-1)(b^2-a^2)^2}\right\}.$$

Proof: Applying Markov's inequality for all $t > 0$, we have,

$$\begin{aligned} P\Big(T_n - E(T_n) \geq \varepsilon\Big) &\leq e^{-t\varepsilon} E \exp\Big\{t \sum_{1\leq i<j\leq n} (X_iX_j - EX_iX_j)\Big\} \\ &= e^{-t\varepsilon} E\Big[\prod_{1\leq i<j\leq n} \exp\Big\{t(X_iX_j - EX_iX_j)\Big\}\Big] \end{aligned}$$

By Lemma 2.1, it is easy to show that the function $\varphi(x_1, x_2, \cdots, x_n) = \exp\Big\{\sum_{1\leq i<j\leq n} x_ix_j\Big\}$ is superadditive. Then properties of NSD random variables and Lemma 2.2, imply that

$$\begin{aligned} P\Big((T_n - E(T_n) \geq \varepsilon\Big) &\leq e^{-t\varepsilon} \prod_{1\leq i<j\leq n} E\Big[\exp\{t(X_iX_j - EX_iX_j)\}\Big] \\ &\leq e^{-t\varepsilon} \prod_{1\leq i<j\leq n} \exp\left\{\frac{1}{8}t^2(b^2-a^2)^2\right\} \\ &= e^{-t\varepsilon} \exp\left\{\frac{\binom{n}{2}}{8}t^2(b^2-a^2)^2\right\} \\ &= \exp\{f(t)\}, \end{aligned}$$

where $f(t) = -\varepsilon t + \frac{n(n-1)}{16}t^2(b^2-a^2)^2$.
Minimizing $f(t)$ we get $t_{\min} = \frac{8\varepsilon}{n(n-1)(b^2-a^2)^2}$ and $f(t_{\min}) = \frac{-4\varepsilon^2}{n(n-1)(b^2-a^2)^2}$.
Replacing in this arguments X_i by $-X_i$ has a similar inequality. These complete the proof.

Corollary 3.1. *Under the assumptions of Theorem 3.1, for every $\varepsilon > 0$,*

$$P\Big(U_n - E(U_n) \geq \varepsilon\Big) \leq \exp\left\{\frac{-\varepsilon^2 n(n-1)}{(b^2-a^2)^2}\right\},$$

where $U_n = \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} X_iX_j$.

Theorem 3.2. *Let $\{X_i; i \geq 1\}$ be a sequence of non-negative NSD random variables and $P(a \leq X_i \leq b) = 1$, $i = 1, 2, \cdots$, and assume that $\{a_{ij}; 1 \leq i < j \leq n\}$ be an array of nonnegative real numbers. Then for all $\varepsilon > 0$,*

$$P\Big(Q_n - E(Q_n) \geq \varepsilon\Big) \leq \exp\left\{\frac{-2\varepsilon^2}{(b^2-a^2)^2 A_n}\right\},$$

where $A_n = \sum_{1\leq i<j\leq n} a_{ij}^2$. **Proof:** Applying, Markov's inequality for all $\varepsilon > 0$ we have,

$$\begin{aligned} P\Big(Q_n - E(Q_n) \geq \varepsilon\Big) &\leq e^{-t\varepsilon} E \exp\Big\{t \sum_{1\leq i<j\leq n} a_{ij}(X_iX_j - EX_iX_j)\Big\} \\ &= e^{-t\varepsilon} E\Big[\prod_{1\leq i<j\leq n} \exp\Big\{ta_{ij}(X_iX_j - EX_iX_j)\Big\}\Big] \end{aligned}$$

By Lemma 2.1, it is easy to show that the function $\psi(x_1, x_2, \cdots, x_n) = \exp\{\sum_{1\leq i<j\leq n} a_{ij}x_ix_j\}$ is superadditive. Then properties of NSD random variables and Lemma 2.2, imply that

$$\begin{aligned} P\Big(Q_n - E(Q_n) \geq \varepsilon\Big) &\leq e^{-t\varepsilon} \prod_{1\leq i<j\leq n} E\exp\{ta_{ij}(X_iX_j - EX_iX_j)\} \\ &\leq e^{-t\varepsilon} \prod_{1\leq i<j\leq n} \exp\left\{\frac{1}{8}t^2(b^2-a^2)^2 a_{ij}^2\right\} \\ &= e^{-t\varepsilon} \exp\left\{\frac{1}{8}t^2(b^2-a^2)^2 A_n\right\} \\ &= \exp\{h(t)\}, \end{aligned}$$

where $h(t) = -\varepsilon t + \frac{t^2}{8}(b^2-a^2)^2 A_n$.
Minimizing $h(t)$ we get $t_{\min} = \frac{4\varepsilon}{(b^2-a^2)^2 A_n}$ and $h(t_{\min}) = \frac{-2\varepsilon^2}{(b^2-a^2)^2 A_n}$.
This completes the proof.

Corollary 3.2. *Under the assumptions of Theorem 3.3, for every $\varepsilon > 0$*

$$P\Big(V_n - E(V_n) \geq \varepsilon\Big) \leq \exp\left\{\frac{-\varepsilon^2 n^2(n-1)^2}{2(b^2-a^2)^2 A_n}\right\},$$

where $V_n = \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} a_ia_jX_iX_j$.

Remark Let $\{X_i; i \geq 1\}$ be a sequence of non-negative NSD random variables with $P(a_i \leq X_i \leq b_i) = 1 \quad (a_i < b_i)$ and $a_i, b_i \in R^+$,$i = 1, 2, ...$, then for all $\varepsilon > 0$ we obtain

$$P\Big(U_n - E(U_n) \geq \varepsilon\Big) \leq \exp\left\{\frac{-\varepsilon^2 n^2(n-1)^2}{2A_n}\right\},$$

and

$$P\Big(V_n - E(V_n) \geq \varepsilon\Big) \leq \exp\left\{\frac{-\varepsilon^2 n^2 (n-1)^2}{2C_n}\right\},$$

where $C_n = \sum_{1\leq i<j\leq n} a_{ij}^2 (b_j b_i - a_j a_i)^2$. and $A_n = \sum_{1\leq i<j\leq n} (b_i b_j - a_i a_j)^2$.
Cuzich et al. (1995), randomizing the sums $\sum_{i<j} X_i X_j$ by products of Rademacher variables prove that

$$\frac{1}{\gamma_n}\sum_{i<j} X_i X_j \overset{a.e.}{\to} 0 \;\Rightarrow \frac{1}{\gamma_n}\sum_{i<j} \varepsilon_i\varepsilon_j X_i X_j \overset{a.e.}{\to} 0 \quad as \;\; n\to\infty,$$

where $\{\varepsilon_i; i \geq 1\}$ is a sequence of independent Rademacher variables independence of $\{X_i; i \geq 1\}$ and $\{\gamma_n; n \geq 1\}$ be a nondecreasing sequence of positive real numbers such that $\gamma_n \to \infty \;\; as \;\; n \to \infty$. In the following theorems, we prove this result using Theorem 3.1 for non-negative NSD uniformly bounded random variables and the sequences of independent Rademacher and Bernoulli variables independence of $\{X_i; i \geq 1\}$.

Theorem 3.3. *Let $\{X_i; i \geq 1\}$ be a sequence of non-negative NSD random variables with $P(a \leq X_i \leq b) = 1$, $i = 1, 2, \cdots$ and assume $\{\varepsilon_i; i \geq 1\}$ be a sequence of independent Rademacher variables independence of $\{X_i; i \geq 1\}$. Let $\{\gamma_n; n \geq 1\}$ be a nondecreasing sequence of positive real numbers, then*

$$P\Big[|\sum_{1\leq i<j\leq n} \varepsilon_i\varepsilon_j X_i X_j| > \gamma_n\varepsilon\Big] \leq \exp\left\{-\frac{4(\gamma_n\varepsilon - b^2\frac{n(n-1)}{2})^2}{n(n-1)(b^2-a^2)^2}\right\}.$$

Proof: Rademacher variables are generally understood as an i.i.d sequence of random variables taking the values -1 and +1 each with probability 1/2. Set $A_n = \{(i,j); \varepsilon_i\varepsilon_j = 1\}$. Since $\{\varepsilon_i; i \geq 1\}$ is a sequence independent of $\{X_i; i \geq 1\}$

we have,

$$\begin{aligned}
P\Big[|\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j X_iX_j|>\gamma_n\varepsilon\Big] &\le P\Big[\sum_{1\le i<j\le n}|\varepsilon_i\varepsilon_j|X_iX_j>\gamma_n\varepsilon\Big]\\
&= P\Big[\sum_{1\le i<j\le n}|\varepsilon_i\varepsilon_j|X_iX_j>\gamma_n\varepsilon|A_n\Big]P[A_n]+\\
&\quad P\Big[\sum_{1\le i<j\le n}|\varepsilon_i\varepsilon_j|X_iX_j>\gamma_n\varepsilon|A_n^c\Big]P[A_n^c]\\
&= P\Big[\sum_{1\le i<j\le n}X_iX_j>\gamma_n\varepsilon\Big]\\
&\le e^{-t\gamma_n\varepsilon}E\Big[\exp(t\sum_{i<j}X_iX_j)\Big]\qquad(\forall\ \ t>0)\\
&\le \exp\Big[-t\gamma_n\varepsilon+(tb^2+\frac{t^2(b^2-a^2)^2}{8})\binom{n}{2}\Big]\\
&\le \exp\left\{-\frac{4(\gamma_n\varepsilon-b^2\frac{n(n-1)}{2})^2}{n(n-1)(b^2-a^2)^2}\right\},
\end{aligned}$$

the third inequality follow from Lemma 2.4 and properties of NSD. This completes the proof.

Theorem 3.4. *Let $\{X_i;i\ge1\}$ be a sequence of non-negative NSD random variables with $P(a\le X_i\le b)=1$, $i=1,2,\cdots$ and assume $\{\varepsilon_i;i\ge1\}$ be a sequence of independent Bernoulli variables independence of $\{X_i;i\ge1\}$. Let $\{\gamma_n;n\ge1\}$ be a nondecreasing sequence of positive real numbers, then*

$$P\Big[|\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j X_iX_j|>\gamma_n\varepsilon\Big]\le\frac{1}{4}P\Big[\sum_{1\le i<j\le n}X_iX_j>\gamma_n\varepsilon\Big].$$

Proof: Set $B_n=\{(i,j);\varepsilon_i\varepsilon_j=1\}$. Since $\{\varepsilon_i;i\ge1\}$ is a sequence independent of $\{X_i;i\ge1\}$ we have,

$$\begin{aligned}
P\Big[|\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j X_iX_j|>\gamma_n\varepsilon\Big] &\le P\Big[\sum_{1\le i<j\le n}|\varepsilon_i\varepsilon_j|X_iX_j>\gamma_n\varepsilon\Big]\\
&= P\Big[\sum_{1\le i<j\le n}|\varepsilon_i\varepsilon_j|X_iX_j>\gamma_n\varepsilon|B_n\Big]P[B_n]\\
&= \frac{1}{4}P\Big[\sum_{1\le i<j\le n}X_iX_j>\gamma_n\varepsilon\Big].
\end{aligned}$$

Corollary 3.3. *Under the assumptions of Theorems 3.5 and 3.6, if*

$$\sum_{n=1}^{\infty}\exp\left\{-\frac{4(\gamma_n\varepsilon-b^2\frac{n(n-1)}{2})^2}{n(n-1)(b^2-a^2)^2}\right\}<\infty,$$

then

$$\lim_{n\to\infty}\frac{1}{\gamma_n}\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j X_iX_j=0 \quad \textit{completely.}$$

In particular, if for all $\varepsilon>0$ *and some* $0<\alpha<1$,

$$\frac{1}{\varepsilon}(n^{\frac{\alpha}{2}}\sqrt{\frac{n(n-1)}{2}}+b^2\frac{n(n-1)}{2})<\gamma_n\le\frac{n(n-1)}{2}.$$

Then

$$\sum_{n=1}^{\infty}\exp\left\{-\frac{4(\gamma_n\varepsilon-b^2\frac{n(n-1)}{2})^2}{n(n-1)(b^2-a^2)^2}\right\}\le\sum_{n=1}^{\infty}\exp(-\frac{2n^{\alpha}}{(b^2-a^2)^2})<\infty,$$

and $\gamma_n\to\infty$, *as* $n\to\infty$. *So,*

$$\lim_{n\to\infty}\frac{1}{\gamma_n}\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j X_iX_j=0 \quad \textit{completely.}$$

4. Examples

Hu (2000) proved a number of well known multivariate distributions possess the NSD property. Among them, multivariate FGM family with margins F_i, $i=1,2,...,n$ for which support of F_i is bounded and Dirichlet distribution satisfy given conditions in section 3. In the following, we describe multivariate FGM family and Dirichlet distribution.

Example 4.1 (FGM family). Let $X_1,X_2,...,X_n$ be random variables with the following joint distribution function;

$$F(x_1,x_2,...,x_n)=\prod_{i=1}^{n}F_i(x_i)[1+\sum_{i<j}\theta_{ij}(1-F_i(x_i))(1-F_j(x_j))], \quad (1)$$

where, the admissible rang of θ_{ij} is such that the density in (1) is non-negative. Also, $(X_1,X_2,...,X_n)$ is NSD if $\theta_{ij}\le 0,\forall\ \ i\ne j$.(for more details on FGM distributions see, Mari and Kotz, 2001).

In particular, if $X_i\sim U(0,1)$, $i=1,2,...,n$, then $E(X_iX_j)=\frac{1}{9}+\frac{\theta_{ij}}{124},\forall i\ne j$. Now, applying Theorem 3.1 we have,

$$P\left[\frac{2}{n(n-1)}\sum_{1\le i<j\le n}(X_iX_j-\frac{1}{9}-\frac{\theta_{ij}}{124})\ge\varepsilon\right]\le e^{-\varepsilon^2 n(n-1)}\ \ \forall\ \ \varepsilon>0.$$

Example 4.2(The Dirichlet distribution). Let $X_1,X_2,...,X_n$ are random variables with density function as

$$f(x_1,x_2,...,x_n)=\frac{\Gamma(\sum_{i=0}^{n}\theta_i)}{\prod_{i=0}^{n}\Gamma(\theta_i)}(1-\sum_{j=0}^{n}x_j)^{\theta_0-1}\prod_{j=1}^{n}x_j^{\theta_j-1},\quad x_j\ge 0,\quad \sum_{j=1}^{n}x_j<1,$$

where the parameter vector $(\theta_0, \theta_1, ..., \theta_n)$ satisfies $\theta_j \geq 1, \quad j = 1, 2, ..., n$. Also, $(X_1, X_2, ..., X_n)$ is NSD and it is obvious $P(0 \leq X_j \leq 1) = 1, \quad j = 1, 2, ..., n$. It is easy to show that for all $i \neq j$, $E(X_i X_j) = \frac{\theta_i \theta_j}{(\theta_0+\theta_i+\theta_j)(\theta_0+\theta_i+\theta_j+1)}$. Therefore, for every $\varepsilon > 0$, we get

$$P\left[\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} (X_i X_j - \frac{1}{9} - \frac{\theta_i \theta_j}{(\theta_0 + \theta_i + \theta_j)(\theta_0 + \theta_i + \theta_j + 1)}) \geq \varepsilon\right] \leq e^{-\varepsilon^2 n(n-1)}.$$

References

1. Christofides, T.C.(1991). Probability inequalities with exponential bounds for U-statistics. *Statist. Probab. Lett.*, **78**, 3294-3297.
2. Christofides, T.C.(1994). A Kolmogorov inequality for U-statistics based on Bernoulli kernels. *J. Statist. Plann. and Inference*, **21**, 357-362.
3. Cuzich, J., Gine, E. and Zinn, J. (1995). Laws of large numbers for quadratic forms, maxima of products and truncated sums of i.i.d. random variables. *The Ann. Probab.*, **23**, 292-333.
4. Eghbal, N., Amini, M. and Bozorgnia, A. (2010). Some maximal inequalities for quadratic forms of negative superadditive dependence random variables. *Statist. Probab. Lett.*, **80**, 587-591.
5. Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *The Ann. Math. Stat.*, **19**, 293-325.
6. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Stat.Assoc.*, **58**, 13-30.
7. Hu, T. (2000). Negatively superadditive dependence of random variables with applications. *Chinese Journal of Applied Probability and Statistics*, **16**, 133-144.
8. Kemperman, J.H.B. (1977). On the FKG-inequalities for measures on a partially ordered space. *Proc. Akad. Wetenschappen*, Ser. A, **80**, 313-331.
9. Mari, D.D. and Kotz, S. (2001). Correlation and dependence. Imperial College Press.
10. Mavrikiou, P.M. (2007). Kolmogorov inequalities for the partial sum of independent Bernoulli random variables. *Statist. Probab. Lett.*, **77**, 1117-1122.
11. Mavrikiou, P.M. (2008). A Kolmogorov inequality for weighted U-statistics. *Statist. Probab. Letters*, **78**, 3294-3297.
12. Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
13. Turner, D.W., Young, D.M. and Seaman, J.W. (1995). A Kolmogorov inequality for the sum of independent Bernoulli random variables with unequal means. *Statist. Probab. Lett.*, **23**, 243-245.
14. Varberg, D.E., 1966. *Convergence of quadratic forms in independent random variables*, Ann. Math. Statist., **37**, 567-575.
15. Whittle, P., 1960. *Bounds for the moments of linear and quadratic forms in independent variables*, Theory Probab. Appl., **5**, 303-305.
16. Whittle, P., 1964. *On the convergence to normality of quadratic forms in independent variables*, Theory Probab. Appl., **9**, 103-109.

17. Young, D.M., Seaman, W.J. and Marco, R.V. (1987). A note on a Kolmogorov inequality. *Statist. Probab. Lett.*, **5**, 217-218.
18. Zhang, C.H., 1996. *Strong law of large numbers for sums of products*, Ann. Probab., **24**, 1589-1615.

An Economy Search Design for 2^m Factorial Experiments

N. Esmailzadeh and H. Talebi

Department of Statistics, University of Isfahan

Since introduction of search design by Srivastava(1975), construction of it has been done by many researchers for 2^m factorial experiments. One of the main goals in construction of the search design is obtaining a design with a highly structured small set of runs. The coherent structure allows for generalization of the searching property for all m. In this paper we obtain a highly structured search design with smaller number of runs than the former ones for all 2^m factorial experiments, $m > 3$.

Keywords: Non-Negligible Effect; Search Designs; Search Linear Model; Bibd.

1. Introduction

Factorial designs are most commonly used designs in the area of experimental designs. In factorial experiments, by hierarchical principal, we assume that the lower order effects are more important than the higher order effects. That is, in a screening stage, we assume that the interaction effects are all negligible. A main effects plan allows us to estimate the general mean and the main effects under the standard linear model assuming no interactions are present. Such an assumption may not be true in reality because there can be a few nonzero interactions present. Consequently, the estimates of the parameters are biased. This motivates the use of search designs for estimating the general mean and main effects as well as searching for identifying the nonzero interactions under the search linear model introduced in Srivastava (1975). Search designs have been constructed based on various statistical models. Several authors have investigated search designs for the main effect plus k plans which are capable of estimating all main effects and identifying and estimating up to k unknown interactions. For example, Shirakura (1991, 1993), Ghosh and Talebi(1993), Mukerjee and Chatterjee (1994), Chatterjee et al.(2001), Shirakura et al.(2002). Consider the following linear model for a 2^m factorial experiment with N runs, $N < 2^m$,

$$y = A_1\xi_1 + A_2\xi_2 + e, \quad Var(e) = \sigma^2 I, \tag{1}$$

where $y(N \times 1)$ is a vector of observations, $A_i(N \times \nu_i)$ are known design matrices and $\xi_i(\nu_i \times 1)$ are vectors of factorial effects for $i = 1, 2$, e is an error random

vector, σ^2 is the error variance and I is the identity matrix. The elements of ξ_1 are unknown parameters. We know that at most k elements of ξ_2 are nonzero but we do not know which elements. The goal is to search for and identify the nonzero elements of ξ_2 and then estimate them along with the elements of ξ_1. Such a model is called a search linear model and the corresponding design matrix is called search design. Let A_{22} be any $(N \times 2k)$ submatrix of A_2. A design is a search design (Srivastava, 1975) if, for every A_{22},

$$rank[A_1; A_{22}] = \nu_1 + 2k. \tag{2}$$

For a 2^m factorial experiment consider the search linear model (1) where ξ_1 consists of the general mean and main effects and ξ_2 is restricted to 2- and 3-factor interactions. Let the matrix $T_1 = [\mathbf{1}'_{\mathbf{m}} : \mathbf{I_m}]$, where $\mathbf{I_m}$ and $\mathbf{1_m}$ are the identity matrix of order m and vector of ones, respectively. Consider a symmetric BIB design with parameters $v = b = 4\lambda - 1, r = k = 2\lambda$ and λ, where $\lambda(\geq 2)$ is a positive integer. Let $T_2(m \times m)$ be its incidence matrix with $m = v = b$. Ghosh and Talebi(1993) showed that the design $T = [T_1 : T_2]$ is a search design with $2m + 1$ runs for $k = 1$. Main effect plans that allow search and estimation of one nonzero element of ξ_2 are called main effect plus one (MEP.l) plans.

Due to decrease cost of running an experiment, it is desired to choose a design with smaller number of runs from set of designs which are considered for a particular purpose. In this paper, we show that in T_1 the run $\mathbf{1}'_{\mathbf{m}}$ is redundant. That is, the reduced design obtained after removing the first row from T do the search problem job with $2m$ runs. In section 2, we present some preliminaries and notations. The main result is given in section 3 for noiseless case, $\sigma^2 = 0$. The noisy case, $\sigma^2 > 0$, will be considered in section 4.

2. Preliminaries and notations

Consider a factorial experiment with m factors each at two levels. A treatment combination is denoted by $(t_1, t_2, ..., t_m)$, where t_i, the level of the $i - th$ factor, is 0 or 1 for $i = 1, ..., m$. Thorough the paper, $\mathbf{0_m}$ denote the vector with all m elements 0 and by $\mathbf{O}_{n\times m}$ we mean the $n \times m$ matrix with all elements 0. Let T be a design with m factors and N treatments(runs) which has been written as an $N \times m$ matrix. We assume that four-factor and higher order interactions are negligible. So, the whole model matrix for T is given by

$$[\mathbf{1}_N; U_1; U_2; U_3]; \tag{3}$$

where $\mathbf{1}_N$ is the column corresponding to the general mean, U_1, U_2 and U_3 are defined as

$$\begin{aligned} U_1 &= [u_1, u_2, ..., u_m] \\ U_2 &= [u_{12}, u_{13}, ..., u_{m-1,m}] \\ U_3 &= [u_{123}, ..., u_{m-2,m-1,m}], \end{aligned} \tag{4}$$

in which u_i for $i = 1, ..., m$ is the column corresponding to the main effect of factor i , u_{ij} $(1 \leq i < j \leq m)$ is the column for the two-factor interaction of factors i and j and u_{ijk} for $(1 \leq i < j < k \leq m)$ is the column for the three-factor interaction of factors i , j and k. Note that

$$\begin{aligned} u_i &= 2\mathbf{t_i} - \mathbf{1}, && i = 1, ..., m, \\ u_{ij} &= u_i * u_j, && 1 \leq i < j \leq m, \\ u_{ijk} &= u_i * u_j * u_k, && 1 \leq i < j < k \leq m, \end{aligned} \tag{5}$$

where $\mathbf{t_i}$ is the $i - th$ column of T and $*$ denotes the Hadamard product. The matrices A_1 and A_2, in (1) are $[\mathbf{1}; U_1]$ and $[U_2; U_3]$, respectively. For the MEP.1 plan (2) is reduced to

$$rank[A_1; A_{22}] = m + 3 \tag{6}$$

Using (5) we define $\mathbf{t_i}$, $\mathbf{t_{ij}}$ and $\mathbf{t_{ijk}}$, according to Chatterjee et.al. (2001), as the following

$$\begin{aligned} &(u_i + \mathbf{1})/2 = \mathbf{t_i}, \\ &(u_{ij} + 2(\mathbf{t_i} + \mathbf{t_j}) - \mathbf{1})/4 = \mathbf{t_i} * \mathbf{t_j} = \mathbf{t_{ij}}, \\ &-(u_{ijk} - 2(\mathbf{t_i} + \mathbf{t_j} + \mathbf{t_k}) + \mathbf{1})/4 = -2\mathbf{t_i} * \mathbf{t_j} * \mathbf{t_k} + \mathbf{t_i} * \mathbf{t_j} + \mathbf{t_i} * \mathbf{t_k} + \mathbf{t_j} * \mathbf{t_k} = \mathbf{t_{ijk}}. \end{aligned} \tag{7}$$

Clearly, an element in vector $\mathbf{t_{ij}}$ is 1 when the corresponding elements of both $\mathbf{t_i}$ and $\mathbf{t_j}$ are 1. It is also can be seen that an element in $\mathbf{t_{ijk}}$ is 1 when at least two of the corresponding elements of $\mathbf{t_i}$, $\mathbf{t_j}$ and $\mathbf{t_k}$ are 1. Implying the elementary column operations in (7) on whole model matrix (3), gives the following matrix

$$[\mathbf{1}; \mathcal{T}_1; \mathcal{T}_2; \mathcal{T}_3]; \tag{8}$$

where $\mathcal{T}_1 = [\mathbf{t_1}, \mathbf{t_2}, ..., \mathbf{t_m}]$, $\mathcal{T}_2 = [\mathbf{t_{12}}, \mathbf{t_{13}}, ..., \mathbf{t_{m-1,m}}]$ and $\mathcal{T}_3 = [\mathbf{t_{123}}, ..., \mathbf{t_{m-2,m-1,m}}]$.

Note that this transformation on design matrix do not change the rank in (6). This implies that we can directly verify condition (6) in (3) using (8).

3. Main result

Consider the matrix

$$T = \begin{bmatrix} \mathbf{I_m} \\ T_2 \end{bmatrix}, \tag{9}$$

where $T_2(m\times m)$ is the incidence matrix of a symmetric $BIBD(4\lambda-1,2\lambda,\lambda)$ with $m=v=b$ and $\lambda(\geq 2)$ is a positive integer. From the properties of the symmetric $BIBD$, it is known that any two columns of T_2 have λ $(1,1),(1,0)$ and $(0,1)$ and $\lambda-1$ $(0,0)$. Note that the matrix in (8) for design T in (9) is

$$\begin{bmatrix} \mathbf{1_m} & \mathbf{I_m} & \mathbf{O_{m\times\binom{m}{2}}} & \mathbf{O_{m\times\binom{m}{3}}} \\ \mathbf{1_m} & T_2 & \mathbf{T}_{22} & \mathbf{T}_{33} \end{bmatrix} \tag{10}$$

where $\mathbf{T}_{22}$ and $\mathbf{T}_{33}$ are the $m\times\binom{m}{2}$ and $m\times\binom{m}{3}$ submatirx of $\mathcal{T}_2$ and $\mathcal{T}_3$ corresponding to T_2, respectively. The following lemmas and their corollaries are useful for the subsequent results.

Lemma 3.1. *Suppose i_1,i_2 and i_3 are three distinct columns of a symmetric $BIBD(4\lambda-1,2\lambda,\lambda)$ incidence matrix. The number of each triple $(1,1,0),(1,0,1)$ and $(0,1,1)$ as rows of (i_1,i_2,i_3) are at least 1.*

Lemma 3.2. *Consider every four columns, say,i_1,i_2,i_3 and i_4, of a symmetric $BIBD(4\lambda-1,2\lambda,\lambda)$ incidence matrix. At least one of the rows $(1,1,0,0)$ or $(0,0,1,1)$ occurs in (i_1,i_2,i_3,i_4).*

Remark 3.1. It is true that both occurs in case of rows $(1,1,1,1)$ or $(0,0,0,0)$ occurs at least once. Note that these two lemmas are direct consequent of lemma 1, 2 and 3 and their remarks of Ghosh and Talebi(1993).

Corollary 3.1. *Any two columns in $\mathbf{T}_{22}$ has at least a copy of set rows $\{(1,0),(0,1),(0,0)\}$.* **Proof**: Any two columns in $\mathbf{T}_{22}$ are corresponding to 2-factor interactions which either have a common or no common factor. Consider the $\mathbf{t_{ij}}$ defined in (7). For the former case the proof is obtained from lemma 3.1. For the later case by lemma 3.2 and noting that any two columns in T_2 contains $\lambda-1$ copies of $(0,0)$ the proof is clear.

Corollary 3.2. *Any two columns, one from $\mathbf{T}_{22}$ and another from $\mathbf{T}_{33}$ have at least a copy of set rows $\{(1,a),(0,1),(0,0)\}$, where a is either 0 or 1.* **Proof**: Suppose the columns in $\mathbf{T}_{22}$ and $\mathbf{T}_{33}$ are denoted by $\mathrm{t}_{i_1j_1}$ and $\mathrm{t}_{i_2j_2k_2}$, where the sets $\{i_1,j_1\}$ and $\{i_2,j_2,k_2\}$ correspond to factors. There exist three possible cases for two sets: *i*) no common factors, *ii*) one common factor, say $j_1=i_2$ and *iii*) two common factors, say $i_1=i_2$ and $j_1=j_2$. For (i), we can always arrange the columns i_1,j_1,i_2,j_2 and k_2 in T_2 as Table 1. There must always be a run with $(i_1,j_1)=(0,a)$ or $(a,0)$ in the $3-th$ block row for $\lambda>2$. Otherwise, in order to have $\lambda-1$ possible $(0,0)$ for (i_1,i_2) and (j_1,i_2) all must occur in the $4-th$ block row. Having λ rows in this block restrict a to be 0 such that the total number of $(0,0)$ for (i_1,j_1) must be at least $\lambda-2$. By similar argument consider all $\lambda-1$ possible $(0,0)$ of columns (j_1,j_2) and (i_1,j_2) in the 2-nd block row and noting that

Table 1. Arranged T_2 in terms of columns i_2 and j_2

i_1	j_1	i_2	j_2	k_2
		$\mathbf{1}_\lambda$	$\mathbf{1}_\lambda$	
		$\mathbf{1}_\lambda$	$\mathbf{0}_\lambda$	
		$\mathbf{0}_{\lambda-1}$	$\mathbf{0}_{\lambda-1}$	
		$\mathbf{0}_\lambda$	$\mathbf{1}_\lambda$	

total number of $(0,0)$ for (i_1, j_1) in these 2 blocks must be at most $\lambda - 1$. It leads to that have $2(\lambda - 2) < \lambda - 1$. That is, $\lambda < 3$ which contradicts the assumption $\lambda > 2$. Therefore, the $3 - th$ row of Table 1 must have some runs 0 for columns i_1 and j_1. So, by having a $(0,0)$ run in the $3 - th$ row of Table 1 for (i_1, j_1) and from lemma 3.1 and this fact that column $\mathrm{t}_{i_1 j_1}$ has λ elements 1, the proof is clear. For $\lambda = 2$ the proof is simple by completing Table 1.
For (ii), by considering columns i_1, j_1 and j_2 in T_2 and lemma 3.1, we obtain the row $(0,1)$ for $(\mathrm{t}_{i_1 j_1}, \mathrm{t}_{i_2 j_2 k_2})$. From rows $(0,0)$ of columns j_1 and j_2, row $(0,0)$ is also obtained. Note that the number of 1 in column $\mathrm{t}_{i_1 j_1}$ is λ. This complete the proof for this case.
Finally, for (iii), by considering the rows $(1,1)$ and $(0,0)$ for columns i_1 and j_1 in T_2 and by lemma 3.1 the proof is clear.

Corollary 3.3. *Any two columns in* $\mathbf{T}_{33}$ *has at least a copy of set rows* $\{(1,a),(b,1),(0,0)\}$ *or* $\{(1,1),(1,0),(0,1)\}$*, where a and b are either 0 or 1, not both 1 simultaneously.* **Proof**: Let two columns in $\mathbf{T}_{33}$ are denoted by $\mathrm{t}_{i_1 j_1 k_1}$ and $\mathrm{t}_{i_2 j_2 k_2}$, where the sets $\{i_1, j_1, k_1\}$ and $\{i_2, j_2, k_2\}$ correspond to factors. The two sets can be common in 0, 1 or 2 elements. Suppose that these two sets are common in two elements, say $j_1 = i_2$ and $k_1 = j_2$. So, in this case we have 4 distinct columns. By considering all possible combinations for columns j_1 and k_1 and lemma 3.2 the proof is clear. Now, suppose that these two sets are common in one element, say $k_1 = i_2$. We arrange the matrix T_2 in terms of columns j_1 and k_1 similar to Table 1 and consider the columns j_1, $k_1(i_2)$, j_2 and k_2. Set rows $(1,a)$ and $(b,1)$ are obtained by lemma 3.1 and lemma 3.2. By the argument similar to Corollary 3.2(i) for four columns j_1, k_1, j_2 and k_2 the row $(0,0)$ is obtained for two columns. This complete the proof. At last, suppose that sets $\{i_1, j_1, k_1\}$ and $\{i_2, j_2, k_2\}$ have no common element. Consider four columns j_1, k_1, i_2 and j_2 . Consider two possible cases whether at least one of the rows $(1,1,1,1)$ or $(0,0,0,0)$ occurs in these columns or not. If it does then by lemma 2 and its remark we have the set rows $\{(0,0),(1,0),(0,1)\}$ or $\{(1,1),(1,0),(0,1)\}$. Otherwise, there exist two possible nonisomorphic rows sets for these 4 columns given in Table 1 of lemma 3 of Ghosh and Talebi(1993). These possible cases give the set rows $\{(0,0),(b,1),(1,a)\}$ and the proof is complete.

Lemma 3.3. *Let W_{11} be an $m \times m$ matrix, W_{12} an $m \times q$ matrix, W_{21} an $n \times m$ matrix, and W_{22} an $n \times q$ matrix. If $rank(W_{11}) = m$, then*

$$rank \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} = m + rank(W_{22} - W_{21}W_{11}^{-1}W_{12}).$$

See page 98 of Harville(1997) for a proof.

Theorem 3.1. *Design T in (9) is an MEP.1 with $N = 2m$ runs.* **Proof**: In order to prove T in (9) is a search design we need to show that condition (6) is valid for (10), from equivalence rank property of (10) with model matrix in (3). That is, for any two distinct columns $\mathbf{t}_{i_1}$ and $\mathbf{t}_{i_2}$ of $[\mathbf{T}_{22}; \mathbf{T}_{33}]$ we need to show that

$$\text{rank} \begin{bmatrix} \mathbf{1}_m & \mathbf{I}_m & \mathbf{0}_m & \mathbf{0}_m \\ \mathbf{1}_m & T_2 & \mathbf{t}_{i_1} & \mathbf{t}_{i_2} \end{bmatrix} = m + 3.$$

Applying Lemma 3.3 by taking $W_{11} = \mathbf{I}_m,\ W_{21} = T_1,\ W_{12} = [\mathbf{1}_m; \mathbf{0}_m; \mathbf{0}_m]$ and $W_{22} = [\mathbf{1}_m; \mathbf{t}_{i_1}; \mathbf{t}_{i_2}]$, the proof is reduced to show that rank$(W_{22} - W_{21}W_{11}^{-1}W_{12})$= rank$(W_{22}) = 3$. There are three possible choices in choosing two columns from $[\mathbf{T}_{22}; \mathbf{T}_{33}]$:

Case i. two columns from $\mathbf{T}_{22}$,
Case ii. one column from $\mathbf{T}_{22}$ and another from $\mathbf{T}_{33}$,
Case iii. two columns from $\mathbf{T}_{33}$.

For the first case Corollary 3.1 guarantees that W_{22} has rank 3, as we always can identify the following sub-matrix of W_{22}

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & a & 1 \end{bmatrix}, \tag{11}$$

whose rank is 3. The proof for cases ii and iii are obtained from Corollaries 3.2 and 3.3, respectively. This completes the proof.

Theorem 3.2. *For $m = 4\lambda - 1 - \upsilon,\ \lambda \geq 2$ and $\upsilon = 1, 2, 3$ consider design $T = \left[\mathbf{I_m} : T_{2(-\upsilon)}\right]$. Then T is an MEP.1 with $N = 2m+\upsilon$ where $T_{2(-\upsilon)}$ is obtained from T_2 in (9) by removing any υ columns.* The proof is clear from Theorem 3.1.

4. Some comparisons in the noisy case

For the noisy case, $\sigma^2 > 0$, condition (2) is still necessary but is not sufficient. Srivastava(1975) proposed a procedure to minimized the sum of square error of rival models to come up with the true model. Shirakura et al.(1996) developed this by considering the probability, $P[SSE(M_0) < SSM(M)]$, where M_0 is the true

model and M is any alternative model, called the search probability(SP). This leaded to the criterion $G(x,\rho) = 1 - \Phi(c_1\rho) - \Phi(c_2\rho) + 2\Phi(c_1\rho)\Phi(c_2\rho)$, where c_1 and c_2 are functions of x. This criterion depends on design through the x and size of the effect through ρ. One can use G for choosing the better design, where the higher value of G is more desired. In this section the G-criterion is used to compare the proposed designs in this paper with the robust designs obtained by Ghosh-Talebi(1993) for $4 \leq m \leq 6$. Figure 1 shows the G-curve for $0 < \rho < 4$. For $m = 4$, there are two isomorphic classes of designs in search probability for both Ghosh-Talebi(GT) and designs in this paper. The graph on the left shows SP curves for $m = 4$. For $m = 5$, there are two classes in SP for GT and one class for designs in this paper, which its SP curve is shown by '.' in the Figure 1. For $m = 6$, there are one class in SP for both designs, which their curves are given in the right side of Figure 1.

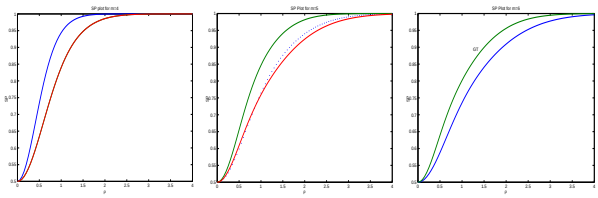

Fig. 1. SP Plot

5. Conclusion

Researchers in their early attempts for constructing the search designs for 2^m factorial experiments made these designs for different values of m, say $m = 4, 5, 6$ and 7, separately. After a while, construction of the designs with search property for general m has been considered. This arise a need for a highly structural set of runs as a design. Ghosh and Talebi(1993) presented such a design for all $m(> 3)$, using the incidence matrix of symmetric $BIBD(4\lambda - 1, 2\lambda, \lambda)$. However, due to reduce the cost of experiments, obtaining search designs with smaller number of runs encourage us to do more challenge. In this paper the new search designs obtained from Ghosh and Talebi(1993) with one less number of runs, yet preserve the searching property of the designs for all $m = 4\lambda - 1$, $\lambda \geq 2$. These designs are very competitive with robust search designs obtained by Ghosh and Talebi(1993) for $m \neq 4\lambda - 1$. The competition is in number of runs for noiseless case and in searching probability capability for noisy case.

References

1. Chatterjee, K.,Deng, L.Y. and Lin, D.k.J., 2001. Two level search design for main-effect plan plus two plan, Metrika, 54, 233-245.
2. Ghosh, S. and Talebi, H., 1993. Main effect plans with additional search property for factorial experiments, J. Statist. Plann. Inference, 36, 367-384.
3. Harville,A. D., 1997. Matrix algebra from a statistician's perspective, Springer-Verlag New York Inc.
4. Mukerjee, R. and Chatterjee, K., 1994. An application of Hadamard matrices for the construction of main effect plus two plans for 2^m factorials. Utilitas Math. 45, 213-218.
5. Shirakura, T., 1991. Main effect plus one or two plans for 2m factorials. J. Statist. Plann. Inference 27, 65-74.
6. Shirakura, T., 1993. Fractional factorial designs of two and three levels. Discrete Math. 116, 99-135.
7. Shirakura, T., Takahashi, T., Srivastava, J.N., 1996. Searching probabilities for nonzero effects in search designs for the noisy case. Ann. Statist. 24 (6), 2560-2568.
8. Shirakura,T., Suetsugu, T. and Tsuji,T., 2002. Construction of main effect plus two plans for 2^m factorials, J. Statist. Plann. Inference, 105 , 405-415.
9. Srivastava, J.N., 1975. Designs for searching non-negligible effects. In: Srivastava, J.N. (Ed.), A Survey of Statistical Design and Linear Model. North-Holland, Amsterdam, pp. 507-519.

Power normal Distribution: Bayesian Estimation

A. Fallah and A. Asgharzadeh

Department of Statistics, University

Department of Statistics, University of Mazandaran

In this paper, we consider a Bayesian approach to estimate the Power normal Distribution parameters. The Bayes estimators are derived for the two unknown parameters and reliability function. The Bayes estimators are derived with respect to conjugate prior for the shape parameter and, discrete prior for the scale parameter of this model. A numerical example and a Monte Carlo simulation study are presented to illustrate the results.

Keywords: Symmetric and asymmetric loss functions; Maximum likelihood estimation; Bayes estimation; Monte Carlo simulation.

1. Introduction

Let Y be a continuous random variable with distribution function $F(y)$. Then we define another random variable X with distribution function $[F(x)]^\alpha$, $\alpha > 0$. This model has been called in the literature Lehmann alternatives. Lehmann (1953) has studied such alternatives to define various non-parametric hypotheses and has computed approximate power of certain rank test using large sample theory. Gupta et al. (1998) studied this model from a reliability point of view discussed the monotonicity of failure rates.In a series of papers Gupta and Kundu (2001a,2002) have developed generalized exponential distribution as an alternative to the gamma, weibull, log-nornal distribution to analyze lifetimes skew data sets. The distribution of X is called a generalized exponential distribution if $F(.)$ is the distribution function. We defined the class of power normal distribution whose distribution function is given by $[\Phi(x)]^\alpha, \alpha > 0$. That is, X is said to have a power normal distribution if its pdf is given by

$$f(x,\alpha) = \alpha[\Phi(x)]^{(\alpha-1)}\phi(x), -\infty < x > \infty, \alpha > 0. \tag{1.1}$$

Note that the density in (1.1) is a weighted normal density with the weight function $[\Phi(x)]^{\alpha-1}$. For $\alpha = 1$ it reduces to the standard normal distribution. The power normal density is a unimodel density which is skewed to the right if $\alpha > 1$ and

to the left if $0 < \alpha < 1$. The power normal distribution, denoted by $PN(x, \alpha)$, has a nice physical interpretation when α is an integer. Let Z be a power normal random variable with the pdf, given in (1.1).Let $X = \sigma Z$. $\sigma > 0$. Then the pdf of X is given by

$$f(x, \alpha, \sigma) = \frac{\alpha}{\sigma}[\Phi(\frac{x}{\sigma})]^{\alpha-1}\phi(\frac{x}{\sigma}). \tag{1.2}$$

and cumulative distribution function (cdf) and reliability function (at some time t)of a two-parameter PN distribution are given, respectively, by

$$F(x, \alpha, \sigma) = [\Phi(\frac{x}{\sigma})]^{\alpha}. \tag{1.3}$$

$$R(t, \alpha, \sigma) = 1 - [\Phi(\frac{t}{\sigma})]^{\alpha}. \tag{1.4}$$

Here σ is the scale parameter and α is the shape parameter. The purpose of this paper is to study the Power normal distribution from the Bayesian point of view. Section 2 contains some preliminaries. In Section 3, we obtain the Bayes estimators of the unknown parameters and reliability function. A numerical example and a Monte Carlo simulation study are given in Section 4.

2. Maximum Likelihood Estimation

Let $X_1, ..., X_n$ be a random sample of size n from (1.2). Then the likelihood function $L(\sigma, \alpha)$ is

$$L(\alpha, \sigma|\mathbf{x}) = \frac{\alpha^n}{\sigma^n}\prod_{i=1}^{n}[\Phi(\frac{x_i}{\sigma})]^{\alpha-1}(\frac{1}{\sqrt{2\pi}})^n \exp(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}x_i^2) \tag{2.1}$$

The log-likelihood function is

$$\begin{aligned} L &= \ln L(\sigma, \alpha|\mathbf{x}) \\ &= n\ln\alpha - n\ln\sigma + (\alpha-1)\sum_{i=1}^{n}\ln[\Phi(\frac{x_i}{\sigma})] - n\ln\sqrt{2\pi} - \frac{-1}{2\sigma^2}\sum_{i=1}^{n}x_i^2. \end{aligned} \tag{2.2}$$

Form (2.2), we obtain the likelihood equations as

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} - \frac{(\alpha-1)}{\sigma^2}\sum_{i=1}^{n}\frac{x_i\phi(\frac{x_i}{\sigma})}{\Phi(\frac{x_i}{\sigma})} + \frac{1}{\sigma^3}\sum_{i=1}^{n}x_i^2 = 0, \tag{2.3}$$

and

$$\frac{\partial L}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n}\ln\Phi(\frac{x_i}{\sigma}) = 0. \tag{2.4}$$

The MLEs $\hat{\alpha}$ and $\hat{\sigma}$ can be obtained by solving the likelihood equations. Solving $(\partial L/\partial\alpha) = 0$ for α gives,

$$\hat{\alpha} = \frac{-n}{\sum_{i=1}^{n} \ln \Phi(\frac{x_i}{\sigma})}., \tag{2.5}$$

where $\hat{\sigma}$ is the solution of

$$-\frac{n}{\sigma} - \frac{1}{\sigma^2}(\frac{-n}{\sum_{i=1}^{n} \ln \Phi \frac{x_i}{\sigma} - 1}) \sum_{i=1}^{n} \frac{x_i \phi(\frac{x_i}{\sigma})}{\Phi(\frac{x_i}{\sigma})} + \frac{1}{\sigma^3} \sum_{i=1}^{n} x_i^2 = 0. \tag{2.6}$$

Newton-Raphson iteration is employed to solve (2.6). The corresponding MLE of the reliability function $R(t)$, is given respectively by (1.4) after replacing α, and σ by their MLE $\hat{\alpha}$, and $\hat{\sigma}$.

2.1. *Loss Function*

In the literature, the most commonly used useful asymmetric loss function is the linear-exponential loss function (LINEX). This loss function was introduced by Varian (1975) and was extensively discussed by Zellner (1986), Soliman (2005) and Soliman et al. (2006). This function rises approximately exponentially on one side of zero, and approximately linearly on the other side.

Under the assumption that the minimal loss occurs at $\phi^* = \phi$, LINEX loss function for $\phi = \phi(\alpha, \sigma)$ can be expressed as

$$L(\Delta) \propto \exp(d\Delta) - d\Delta - 1, \quad d \neq 0, \tag{2.7}$$

where $\Delta = (\phi^* - \phi), \phi^*$ is an estimate of ϕ. The sign and magnitude of the shape parameter d represents the direction and degree of symmetry, respectively. (If $d > 0$, the overestimation is more serious than underestimation, and vice-versa.) For d close to zero, the LINEX loss is approximately SEL and there fore almost symmetric. The posterior expectation of the LINEX loss function (2.7) is

$$E_\phi[L(\phi^* - \phi)] \propto \exp(d\phi^*) E_\phi[\exp(-d\phi)] - d(\phi^* - E_\phi(\phi)) - 1, \tag{2.8}$$

where $E_\phi(\cdot)$ denotes the posterior expectation with respect to the posterior density of ϕ. The Bayes estimator of ϕ, denoted by ϕ^*_{BL} under the LINEX loss function is the value ϕ^* which minimizes (2.8). It is

$$\phi^*_{BL} = -\frac{1}{d} \ln\{E_\phi[\exp(-d\phi)]\}, \tag{2.9}$$

provided that the expectation $E_\phi[\exp(-d\phi)]$ exists and is finite. The problem of choosing the value of the parameter d is discussed in Calabria and Pulcini (1996).

Another useful asymmetric loss function is the Entropy loss

$$L(\phi^* - \phi) \propto \left(\frac{\phi^*}{\phi}\right)^1 - \log\left(\frac{\phi^*}{\phi}\right) - 1, \tag{2.10}$$

whose minimum occurs at $\phi^* = \phi$. The Bayes estimate ϕ^*_{BG} of ϕ under Entropy loss (2.10) is

$$\phi^*_{BG} = \left(E_\phi(\phi^{-1})\right)^{-1}, \tag{2.11}$$

provided that $E_\phi(\phi^{-1})$ exists and is finite.

3. Bayes Estimation

Under the assumption that both the parameters α and σ are unknown, specifying a general joint prior for α and σ leads to computational complexities for the Bayes estimates. To solve this problem and simplify the Bayesian analysis, we consider the method advocated by Soland (1969). In this method, we use a conjugate continuous-discrete joint prior distribution for the parameters α and σ. The continuous component of this distribution is related to α and the discrete one is related to σ. This method has been further used by Soliman (2005) and Soliman et al. (2006) who have discussed Bayesian analysis for Weibull and Burr-XII distributions, respectively.

We assume that the scale parameter σ is restricted to a finite number of values $\sigma_1, \sigma_2, \cdots, \sigma_N$ with prior probabilities $\eta_1, \eta_2, \cdots, \eta_N$ respectively, where $0 \leq \eta_j \leq 1$, and $\sum_{j=1}^{N} \eta_j = 1$, i.e.

$$\pi(\sigma_j) = p_r(\sigma = \sigma_j) = \eta_j, \qquad j = 1, 2, ..., N.$$

Further, suppose that conditional upon $\sigma = \sigma_j,\ j = 1, 2, ..., N$, α has a natural conjugate gamma prior with parameters a_j and b_j

$$\pi(\alpha|\sigma_j) = \frac{b_j^{a_j} \alpha^{a_j - 1} e^{-\alpha b_j}}{\Gamma(a_j)}, \qquad a_j, b_j, \alpha > 0. \tag{3.1}$$

Combining the likelihood function in (2.1) and prior pdf (3.1), we obtain the conditional posterior pdf of α given $\sigma = \sigma_j$ as

$$\pi^*(\alpha|\sigma_j; \mathbf{x}) = \frac{B_j^{A_j} \alpha^{A_j - 1}}{\Gamma(A_j)} \exp(-\alpha B_j), \quad A_j, B_j, \alpha > 0, \tag{3.2}$$

where

$$A_j = n + a_j \quad , B_j = b_j - \sum_{i=1}^{n} \ln \Phi(\frac{x_i}{\sigma_j}). \tag{3.3}$$

The joint prior density $\pi(\alpha, \sigma_j)$ can be obtained by multiplying $\pi(\alpha|\sigma_j)$ by $\pi(\sigma_j)$. Now, by multiplying the likelihood function by the joint prior density, the joint

posterior density of α and σ_j is

$$\pi^*(\alpha,\sigma_j|\mathbf{x}) = \frac{b_j^{a_j} u_j\ \eta_j\ \alpha^{A_j-1} e^{-\alpha[b_j-\sum_{i=1}^n \ln\Phi(\frac{x_i}{\sigma_j})]}}{G\ \Gamma(a_j)\sigma_j^n}, \tag{3.4}$$

and the marginal posterior probability of σ_j is

$$p_j = p_r(\sigma=\sigma_j|\mathbf{x}_j) = \frac{b_j^{a_j}\eta_j u_j \Gamma(A_j)}{G\Gamma(a_j)\sigma_j^n B_j^{A_j}}, \tag{3.5}$$

where

$$G = \sum_{j=1}^{N} \frac{u_j b_j^{a_j}\eta_j\Gamma(A_j)}{\sigma_j^n\Gamma(a_j)B_j^{A_j}}$$

and

$$u_j = \prod_{i=1}^{n} \frac{\phi(\frac{x_i}{\sigma_j})}{\Phi(\frac{x_i}{\sigma_j})}. \tag{3.6}$$

Under a squared error loss function, the usual estimate of a parameter is the posterior mean. Thus, Bayes estimates of the parameters and reliability function are obtained by using (3.2) and (3.5). The Bayes estimates $\tilde{\alpha}_{BS}$, and $\tilde{\sigma}_{BS}$ of parameters α, and σ are

$$\tilde{\alpha}_{BS} = \sum_{j=1}^{N} p_j \frac{(n+a_j)}{[b_j - \sum_{i=1}^n \ln\Phi(\frac{x_i}{\sigma_j})]}. \tag{3.7}$$

and

$$\tilde{\sigma}_{BS} = \sum_{j=1}^{N} p_j\sigma_j \tag{3.8}$$

The Bayes estimate, $\tilde{R}_{BS}(t)$, of the reliability function $R\equiv R(t)$, is

$$\begin{aligned}\tilde{R}_{BS}(t) &= \sum_{j=1}^{N} p_j \int_0^{\infty} \left[1-[\Phi(\frac{t}{\sigma_j})]^{\alpha}\right]\pi^*(\alpha|\sigma_j,\mathbf{x}_j)d\alpha \\ &= \sum_{j=1}^{N} p_j \left(1-\left[1-\frac{\ln(\frac{t}{\sigma_j})}{[b_j-\sum_{i=1}^n \ln\Phi(\frac{x_i}{\sigma_j})]}\right]^{-A_j}\right).\end{aligned} \tag{3.9}$$

Under the LINEX loss function (2.7), the Bayes estimates $\tilde{\alpha}_{BL}$, and $\tilde{\sigma}_{BL}$ of parameters α, and σ are

$$\tilde{\alpha}_{BL} = \frac{-1}{d}\log\left[\sum_{j=1}^{N} p_j\left[1+\frac{d}{(b_j-\sum_{i=1}^n \ln\Phi(\frac{x_i}{\sigma_j}))}\right]^{-A_j}\right]. \tag{3.10}$$

and

$$\tilde{\sigma}_{BL} = \frac{-1}{d} \log \sum_{j=1}^{N} p_j e^{-d\sigma_j}. \tag{3.11}$$

similarly, the Bayes estimator for the reliability function $R(t)$ is given by

$$\tilde{R}(t)_{BL} = \frac{-1}{d} \log \left[\sum_{j=1}^{N} \sum_{s=0}^{\infty} p_j \frac{e^{-d} d^s}{s!} \left[1 - \frac{s \log \Phi(\frac{t}{\sigma_j})}{[b_j - \sum_{i=1}^{n} \ln \Phi(\frac{x_i}{\sigma_j})]} \right]^{-A_j} \right]. \tag{3.12}$$

Under the entropy loss function (2.10), the Bayes estimates $\tilde{\alpha}_{BG}$, and $\tilde{\sigma}_{BG}$, of parameters α, and σ, are

$$\tilde{\alpha}_{BG} = \left[\sum_{j=1}^{N} \frac{p_j (b_j - \sum_{i=1}^{n} \ln \Phi(\frac{x_i}{\sigma_j}))}{(n + a_j - 1)!} \right]^{-1}. \tag{3.13}$$

and

$$\tilde{\sigma}_{BG} = \left[\sum_{j=1}^{N} p_j \sigma_j^{-1} \right]^{-1}. \tag{3.14}$$

Similarly, the Bayes estimator for the reliability function $R(t)$ is given by

$$\tilde{R}(t)_{BG} = \left[\sum_{j=1}^{N} \sum_{s=0}^{\infty} p_j \left[1 - \frac{s \log \Phi(\frac{t}{\sigma_j})}{(b_j - \sum_{i=1}^{n} \ln \Phi(\frac{x_i}{\sigma_j}))} \right]^{-A_j} \right]^{-1}. \tag{3.15}$$

To implement the calculations in this section, it is first necessary to elicit the values of (σ_j, η_j) and the hyper parameters (a_j, b_j) in the conjugate prior (3.6), for $j = 1, 2, \cdots, N$. The hyper parameters (a_j, b_j) can be obtained based on the expected value of the reliability function $R(t)$ conditional on $\sigma = \sigma_j$, which is given using (1.4) and (3.1) by

$$\begin{aligned} E_{\alpha|\sigma_j}[R(t)|\sigma = \sigma_j] &= \int_0^{\infty} \left[1 - [\Phi(\frac{t}{\sigma_j})]^{\alpha} \right] \frac{b_j^{a_j} \alpha^{a_j - 1}}{\Gamma(a_j)} e^{-\alpha b_j} d\alpha \\ &= 1 - \left[1 - \frac{\ln[\Phi(\frac{t}{\sigma_j})]}{b_j} \right]^{-a_j}. \end{aligned} \tag{3.16}$$

Now, suppose that prior beliefs about the lifetime distribution enable one to specify two values $(R(t_1), t_1), (R(t_2), t_2)$. Thus, for these two prior values $R(t = t_1)$ and $R(t = t_2)$, the values of a_j and b_j for each value σ_j, can be obtained numerically from (3.16). If there are no prior beliefs, the non parametric procedure

$$\tilde{R}(t = x_{(i)}) = 1 - \frac{i}{n+1},$$

can be used to estimate the reliability function $R(t)$.

3.1. *Numerical Example.*

For given values of $\alpha = 2, \sigma = 1$, a sample of size $n = 10$ is generated from the power normal distribution. This sample is:

```
0.0069 1.1519 0.1334 -0.3368 0.5271 0.8132 0.8037 0.4316 1.085
0.2718
```

The MLE of α, and σ, using a New-Raphson method when solving (2.5), and (2.6), are obtained as $\widehat{\alpha} = 1.92$ and $\widehat{\sigma} = 0.7805$. Substituting $\widehat{\alpha}$ and $\widehat{\sigma}$ into (1.4), we obtain MLE of the reliability function at $t = 2$ as $\widehat{R}(2) = 1$. To obtain Bayes estimates, it is first necessary to elicit the values of (σ_j, η_j) and the hyper parameters (a_j, b_j) in the conjugate prior (3.1), for $j = 1, 2, \cdots, N$. These values are derived by the following steps:
1. Based on observations, we estimate two values of the reliability function as

$$\tilde{R}(t = 0.0069) = 1 - \frac{i}{n+1} = 1 - \frac{2}{11} = 0.81$$

and

$$\tilde{R}(t = 0.2718) = 1 - \frac{i}{n+1} = 1 - \frac{4}{11} = 0.63,$$

2. Since the MLE of σ is $\widehat{\sigma} = 0.7805$, we assume that σ_j takes the values $0.5(0.1)1.4$, each with probability 0.1.
3. The two prior values obtained in step 1 are substituted into (3.16), where a_j and b_j are solved numerically for each given $\sigma_j, j = 1, 2, \cdots, 10$, using the Newton-Raphson method. Table 1 gives the values of the hyper parameters and the posterior probabilities derived for each σ_j. The MLEs $(\cdot)_{ML}$, and the Bayes estimates $((\cdot)_{BG}, (\cdot)_{BS}, (\cdot)_{BL})$ of α, σ and $R(t)$ computed and the results are displayed in Table 2.

Table 1: prior information, Hyper parameter

values and the posterior probabilities

j	1	2	3	4	5
σ_j	0.5	0.6	0.7	0.8	0.9
η_j	0.1	0.1	0.1	0.1	0.1
a_j	2	0.5012	0.1779	0.1458	$1.066\ e^{-001}$
b_j	3	0.0144	0.0005	0.00006	$1.60e^{-006}$
u_j	1630.67	598.51	304.37	187.09	129.64
p_j	$9.74e^{-001}$	$2.48e^{-002}$	$7.32e^{-004}$	$9.12e^{-005}$	$1.37e^{-005}$
j	6	7	8	9	10
σ_j	1	1.1	1.2	1.3	1.4
η_j	0.1	0.1	0.1	0.1	0.1
a_j	$1.064e^{-001}$	1.062	$1.060e^{-001}$	$1.060e^{-001}$	$1.058e^{-001}$
b_j	$1.58e^{-006}$	$1.49e^{-006}$	$1.31e^{-006}$	e^{-006}	$1.26e^{-006}$
u_j	97.35	77.37	64.09	54.76	47.93
p_j	$3.60e^{-006}$	$1.09e^{-006}$	$3.78e^{-007}$	$1.44e^{-006}$	$6.007\ e^{-008}$

Table 2: Estimators of α, σ and $R(t)$ with $t = 2$

	$(\cdot)_{ML}$	$(\cdot)_{BG}$	$(\cdot)_{BS}$	$(\cdot)_{BL}$		
				$d = -0.5$	$d = 0.5$	$d = 1$
α	1.9200	1.2858	1.3648	1.4852	1.3382	1.3136
σ	0.7805	0.5044	0.5058	0.5076	0.5055	0.5053
$R(t)$	0.9846	0.9930	0.9996	0.9997	0.9996	0.9995

References

1. Calabria, R. Pulcini, G. (1996). Point estimation under asymmetric loss function for left truncated exponential samples. *Commun. Statist. Theory Methods*, 25(3): 585-600.
2. Lehmann, E. (1953). The power of rank tests. *Ann Math Stat*, 24:28-43
3. Gupta, R.C. Gupta, P.L. Gupta, R.D. (1998). Modeling failure time data by lehmann alternatives. *Commun Stat*, 27:887-904
4. Gupta, R.D. kundu, D. (2001). Exponentiated exponential family; an alternative to gamma and weibull distributions. *Biometrical journal*, 43(1):117-130
5. Gupta, R.D. kundu, D. (2002). Generalized exponential distribution:statistical inferences. *J. Stat. Theory*. 101-118
6. Soland; R. M. (1969). Bayesian analysis of the Weibull process with unknown scale and shape parameters. *IEEE Trans. Reliab*, vol R-18: 181-184.
7. Soliman, A. A. (2005). Estimation of parameters of life from progressively censored data using Burr-XII model. *IEEE Trans. Reliab*, 54(1): 34-42.
8. Soliman , A. A; Abd Ellah, A. H., Sultan, K. S. (2006). Comparison of estimates using record statistics from Weibull model: Bayesian and non-Bayesian approachesE. *Comput. Statist. Data Anal*, 51(3): 2065-2077.
9. Varian. H. R. (1975). A Bayesian Approach to Real Estate Assessment. North Holland, Amsterdam, 195-208.

10. Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss function., *J. Amer. Statist. Assoc*, 81: 446-451.

Estimation of Parametric Functions for Discrete Distributions Generated by Cauchy Stable Law

D. Farbod and K. V. Gasparian

Department of Statistics, Islamic Azad University - Mashhad Branch

Yerevan State University, Yerevan

In this paper, we consider some useful parametric functions for discrete distributions generated by stable law. Then, using the asymptotic properties of the maximum likelihood (ML) estimators of the scale parameter γ, we conclude related asymptotic behavior results for such parametric functions.

Keywords: Asymptotic properties, Cauchy stable Law, ML, Stable Laws

1. Introduction

Some well-known discrete distributions are widely used in large-scale biomolecular systems. But the diversity of such systems requires to generate new ones satisfying in empirical statistical properties (see [1]). Based on the empirical facts in large-scale biomolecular systems, Astola and Danielian [1, 2] considered the families of unimodal frequency decreasing distributions with right-side asymmetry (as many *Stable Laws*) which are log-downword convex.

The *Stable Laws* form a rich class of probability distributions allowing skewness, heavy tails and have many "useful" mathematical properties. The notion of *Stable Laws* was introduced by Paul Levy in the 1920s. For more details on this see [14].

Definition 1. ([12], p.7) *Non-degenerate random variable X is stable if and only if for all $n \geq 2$, there is constant $d_n \in \mathbb{R}$ such that*

$$X_1 + X_2 + ... + X_n \stackrel{d}{=} n^{1/\alpha} X + d_n,$$

where $X_1, X_2, ..., X_n$ are independent, identical copies of X and $\alpha \in (0, 2]$. (The symbol $\stackrel{d}{=}$ means equality in distribution).

Any *stable* density $s(x;\alpha,\beta,\gamma,\delta)$ depends on the following parameters: $\alpha \in (0,2]$ - is the index of stability, $\beta \in [-1,1]$ - the skewness parameter, $\gamma \in (0,\infty)$ - the scale parameter, and $\delta \in \mathbb{R}$ - the location parameter. If $\alpha = 1$ and $\beta = 0$, then the distribution is called *Cauchy* stable distribution which has the following density (see [6], [14]):

$$s(x;\gamma,\delta) = s(x;1,0,\gamma,\delta) = \frac{\gamma}{2(\frac{\pi^2}{4}\gamma^2 + (x-\delta)^2)}, \quad -\infty < x < \infty. \tag{1}$$

In the present paper we suppose that

$$\Theta = \{(\gamma,\delta) : \ 0 < \gamma < \infty, \ \delta = 0\},$$

and also G is an arbitrary open subset of Θ whose closure $\overline{G}$ is also contained in Θ.

Let us now consider the following discrete distribution (see [6]):

$$g(x;\gamma) = c_\gamma^{-1} \cdot s(x;\gamma) \qquad x = 0,1,2,..., \tag{2}$$

where $c_\gamma = \sum_{y=0}^{\infty} s(y;\gamma)$.

The reminder of the paper is organized as follows. In Section 2 we give the regularity conditions (RC) for the model (2), under which the well-known asymptotic properties of the ML estimator of the scale parameter γ are met (see [6]). The main results of the paper are given in Section 3.

2. ML Estimators

Supposing $X^n = (X_1,...,X_n)$, with realization $x^n = (x_1,...,x_n)$, is a sample from (2). We consider the following RC:

RC-conditions:

1. $\Theta \subset \mathbb{R}$ is a compact subset of $\mathbb{R}$;

2. $g(x;\gamma_1) \neq g(x;\gamma_2)$ for all $\gamma_1 \neq \gamma_2$, $\gamma_1,\gamma_2 \in \Theta$;

3. probability distributions, say $\mathbb{P}_\gamma$, have a common support, that is the set

$$supp\ \mathbb{P}_\gamma = \{x : g(x;\gamma) > 0\}$$

does not depend from $\gamma \in \Theta$;

4. the function $\ln g(x;\gamma)$ is twice continuously differentiable by γ for all $x = 0, 1, 2, ...$, moreover, there exists a function $D(x)$ for which

$$|\frac{\partial^2 \ln g(x;\gamma)}{\partial \gamma^2}| \leq D(x),$$

and

$$E_\gamma(D(X_1)) < \infty;$$

5. for the true value $\gamma_0 \in \overline{G}$,

$$E_{\gamma_0}[\sup_{\gamma \in \Theta - \overline{G}} \ln \frac{L(X^n;\gamma)}{L(X^n;\gamma_0)}] < \infty,$$

where $L(X^n;\gamma) = \prod_{i=1}^n g(X_i;\gamma)$ is likelihood function;

6. for all $\gamma \in \Theta$ the Fisher's information measure $I(\gamma)$, contains in observation X_1, satisfies the following condition

$$0 < I(\gamma) = E_\gamma[\frac{\partial \ln g(X_1;\gamma)}{\partial \gamma}]^2 = -E_\gamma[\frac{\partial^2 \ln g(x;\gamma)}{\partial \gamma^2}] < \infty,$$

and is continuous at γ.

It was proved in [6], that the RC-conditions are met for the ML estimator of the scale parameter γ of the model (2). Under satisfying this conditions, the following Theorem takes place (see [4], [11]):

Theorem 1. *Assume that the RC-conditions are fulfilled and $\gamma_0 \in G$. Then, the likelihood equation*

$$\frac{\partial \ln L(x^n;\gamma)}{\partial \gamma} = 0, \tag{3}$$

has a unique solution $\widehat{\gamma}_n = \widehat{\gamma}(X^n)$ in G. This solution is a ML estimator and has the following properties:

(a). strong consistency, asymptotic normality and asymptotic efficiency, i.e. as $n \longrightarrow \infty$ then:

$$\widehat{\omega}_n \equiv \sqrt{n}(\widehat{\gamma}_n - \gamma) \xrightarrow{d} \eta \in N(0, I^{-1}(\gamma)).$$

(b). convergence of moments, that is

$$E_\gamma \widehat{\omega}_n^k \longrightarrow E_\gamma \eta^k, \ \forall \ k \geq 1. \tag{4}$$

Remark 1. *From* (4) *when* $k = 1$ *the property of asymptotic unbiasedness also satisfies, i.e.*

$$E_\gamma \widehat{\gamma}_n = \gamma + o(n^{-\frac{1}{2}}).$$

(c). *If* $f(t)$ *is a differentiable function on* $\mathbb{R}$ *such that* $f'(t) \neq 0$, *then*

$$\sqrt{n}(f(\widehat{\gamma}_n) - f(\gamma)) \xrightarrow{d} \eta \in N(0, \frac{[f^{'}(\gamma)]^2}{I(\gamma)}). \tag{5}$$

Remark 3. *From* $k = 2$ *in (5), there take place*

$$E_\gamma(\widehat{\gamma}_n - \gamma)^2 = \frac{1 + o(1)}{n\ I(\gamma)}.$$

The relation (5) *can also be represented as follows*

$$E_\gamma(f(\widehat{\gamma}_n) - f(\gamma))^2 = \frac{[f^{'}(\gamma)]^2}{n\ I(\gamma)} \cdot (1 + o(1)).$$

3. Estimation of Some Parametric Functions

Let us denote

$$s^{'}(x;\gamma) = \frac{\partial s(x;\gamma)}{\partial \gamma}, \quad (c_\gamma)^{'} = \frac{\partial c_\gamma}{\partial \gamma}.$$

Lemma 1. *The Fisher's information measure* $I(\gamma)$ *of the model* (2) *can be represented as follows*

$$I(\gamma) = Var_\gamma\{U[s(X_1;\gamma)]\},$$

where $U[s(X_1;\gamma)] = [\ln s(x;\gamma)]^{'} = \frac{s^{'}(x;\gamma)}{s(x;\gamma)}$ *is contribution function of* X_1 *for the Cauchy stable law* (1).

Proof. It is easily seen that

$$U[g(x,\gamma)] = \frac{s'(x,\gamma)}{s(x,\gamma)} - \frac{(c_\gamma)'}{c_\gamma},$$

and

$$E_\gamma[U(s(X_1,\gamma))] = U(c_\gamma).$$

We have

$$I(\gamma) = E_\gamma U^2[g(X_1,\gamma)] = E_\gamma U^2[s(X_1,\gamma)] - (\frac{(c_\gamma)'}{c_\gamma})^2 = Var_\gamma\{U[s(X_1,\gamma)]\}.$$

The finiteness of $I(\gamma)$ is obvious. The proof of Lemma 1 is complete.

With the help of Theorem 1 and Lemma 1 we find now the estimators of some useful parametric functions of the model (2). Here they are:

1. $\sqrt{n}(I(\widehat{\gamma}_n) - I(\gamma)) \xrightarrow{d} N(0, \frac{[I^{'}(\gamma)]^2}{I(\gamma)})$,

for $\tau(\gamma) = I(\gamma)$.

2. $\sqrt{n}(g(x;\widehat{\gamma}_n) - g(x;\gamma)) \xrightarrow{d} N(0, \frac{[g^{'}(x;\gamma)]^2}{I(\gamma)})$,

for $\tau_x(\gamma) = g(x;\gamma) = \frac{s(x;\gamma)}{c_\gamma}$, $x \in \mathbb{N} \cup \{0\}$, where

$$g^{'}(x;\gamma) = U[s(X_1;\gamma)] - g(x;\gamma) \cdot E_\gamma U[s(X_1;\gamma)].$$

3. for $\tau_t(\gamma) = \overline{F}_\gamma(t) \equiv 1 - F_\gamma(t) = \sum_{x=t}^{\infty} g(x;\gamma)$, $t \in (0,\infty)$,

$$\sqrt{n}(\overline{F}_{\widehat{\gamma}_n}(t) - \overline{F}_\gamma(t)) \xrightarrow{d} N(0, \frac{[\overline{F}^{'}_\gamma(t)]^2}{I(\gamma)}),$$

where

$$\overline{F}^{'}_\gamma(t) = (1_{X_1 \geq t} - \overline{F}_\gamma(t)) \cdot E_\gamma U[s(X_1;\gamma)].$$

4. for $\tau(\gamma) = c_\gamma = \sum_{y=0}^{\infty} s(y;\gamma)$,

$$\sqrt{n}(c_{\widehat{\gamma}_n} - c_\gamma) \xrightarrow{d} N(0, \frac{[(c_\gamma)^{'}]^2}{I(\gamma)}),$$

where $(c_\gamma)^{'} = c_\gamma \cdot E_\gamma U[s(X_1;\gamma)]$.

5. for $\tau(\gamma) = \frac{(c_\gamma)^{'}}{c_\gamma} = U(c_\gamma)$

$$\sqrt{n}(U(c_{\widehat{\gamma}_n}) - U(c_\gamma)) \xrightarrow{d} N(0, \frac{[U'(c_\gamma)]^2}{I(\gamma)}),$$

where $U'(c_\gamma) = E_\gamma U^{'}[s(X_1;\gamma)]$.

References

1. J. Astola and E. Danielian, *Frequency distributions in biomolecular systems and growing networks*, TICSP Series No. 31, Tampere, Finland, 2006.
2. J. Astola and E. Danielian, *Regualrly varying skewed distributions generated by birth-death process,* TICSP Series, No. 27, Tampere, Finland, 2004.
3. J. Astola, K. V. Gasparian and E. Danielian, The maximum likelihood estimators for distributions with moderate growth, *Proceedings of CSIT Conference*, Sept. 24-28, Yerevan, Armenia, (2007), pp. 91-94.

4. A. A. Borovkov, *Mathematical Statistics*, Gordon and Breach Science Publishers, 1998 (translated from the original Russian ed.).
5. W. H. DuMouchel, On the asymptotic normality of the maximum likelihood estimate when sampling from a stable distribution, *Ann. Statist.*, **1**, (1973), pp. 948-957.
6. D. Farbod and K .V. Gasparian, Asymptotic properties of the maximum likelihood estimator for some discrete distributions generated by Cauchy stable law, *"STATISTICA"* (Bologna), Italy, Vol. 68, No. 3-4, (2008).
7. D. Farbod and K. V. Gasparian, The asymptotic properties of some discrete distributions generated by standard skwewed stable laws, *Proceedings of the* 9_{th} *Iranian Statistical Conference*, Agust 20-22, Isfahan, Iran, (2008), pp. 138-144.
8. D. Farbod, Asymptotic Properties of Parameters Estimators for Stable Laws Discretizations and Their Numerical Approximations, Ph.D Thesis, Yerevan State University, Yerevan, Armenia, (Sept. 2009), 80 pp.
9. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, 2nd ed., John Wiley and Sons, 1971.
10. H. Fofack and J. P. Nolan, Modes and other characteristics of stable distributions, *Extremes*, **2**, (1999), 39-58.
11. E. L. Lehmann, *Theory of Point Estimation*, Wiley, New York, 1983.
12. J. P. Nolan, *Stable Distributions - Models for Heavy Tailed Data*, Boston: Birkhauser, In progress, chapter 1 online at academic 2, American.edu/∼ Jpnolan, 2008.
13. G. Samorodnitsky and M. S. Taqqu, *Stable non-Gaussian random processes, stochastic models with infinite variance*, Chapman & Hall, New York, 1994.
14. V. M. Zolotarev, *One-dimensional stable distributions*, Volume 65 of translations of mathematical monographs, American Mathematical Society, 1986 (translated from the original 1983 Russian ed.).

Modelling RLC Electrical Circuits By Stochastic Differential Equations

R. Farnoosh and P. Nabati

Department of Applied Mathematics, Iran University of Science and Technology

The main aim of this study is to solve the RLC electrical circuit when one or some of the coefficients are effected by noise. For this purpose the deterministic model is replaced by stochastic model and then this model will be solved analytically and numerically. Computer programs in MATLAB are used to generate numerical simulations and their graphical representations.

Keywords: Stochastic Differential Equation, White noise, Simulation, Electrical Circuits.

1. Introduction

In this work we motivate the use of second order stochastic differential equations (SDEs) in RLC electrical circuits. In recent decades important advance have been made in modelling based on SDEs. These have been applied in many scientific fields such as engineering, environmental modelling, biology and etc. The main part of stochastic calculus are the Ito calculus and Stratonovich. We consider SDEs of kind Ito in this paper. A general SDE is given by

$$dX(t) = f(t, X(t))dt + g(t, X(t))dW(t), \tag{1}$$

where $f(t, X(t))$ and $g(t, X(t))$ are drift and diffusion term, respectively, and $W(t)$ is a Wiener process. Sometimes SDEs can not be solved analytically. It is extremely difficult and even not impossible that we solve their exact solutions in most cases, therefore we can obtain their approximate numerical solution through numerical simulation. This paper focus on the Euler-Maruyama scheme that has a strong order of accuracy. The outline of this paper is as follows. In section 2, the problem formulation in deterministic and stochastic models will be presented. Section 3 describe the method of analytical solution for solving two dimensional SDEs. Finally numerical solutions are obtained using the Euler scheme in section 4.

2. Problem formulation

Any electrical circuit consists of resistor(R), capacitor(C) and inductor(L). These circuit elements can be combined to form an electrical circuits in four distinct ways: the RC, RL, LC and RLC circuits. Then, an (RLC) circuit, is an electrical circuit composed of resistor, capacitor and inductor driven by a voltage or current source. The charge $Q(t)$ at time t at a fixed point in an electrical circuit according Kirchhoff's law satisfies the differential equation [8]

$$-V(t) + RI(t) + L\frac{dI}{dt} + \frac{1}{C}\int I(t)dt = 0, \tag{2}$$

where $I(t) = \frac{dQ(t)}{dt}$ then

$$L\ddot{Q}(t) + R\dot{Q}(t) + \frac{1}{C}Q(t) = V(t),\ Q(0) = Q_0,\ \dot{Q}(0) = I_0, \tag{3}$$

where $V(t)$ is the potential source at time t.
Now we may have a situation where some of the coefficients, say $V(t)$, are not deterministic but of the form

$$V^*(t) = V(t) +' noise'.$$

Observation indicated that the noise can be described as a multiple of the so called "white noise process" denoted by $W(t)$, we get the following equation,

$$L\ddot{Q}(t) + R\dot{Q}(t) + \frac{1}{C}Q(t) = V(t) + \alpha W_t,\ Q(0) = Q_0, \dot{Q}(0) = I_0, \tag{4}$$

where α is the intensity of noise. This is a second order stochastic differential equation.

3. The Analytical Solution

In order to solve analytically equation(4), we introduce the vector
$X = \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} Q(t) \\ \dot{Q}(t) \end{pmatrix}$ and obtain

$$\begin{cases} dX_1(t) = X_2(t)dt \\ LdX_2(t) = (-RX_2(t) - \frac{1}{C}X_1(t) + V(t))dt + \alpha dB_t, \end{cases} \tag{5}$$

or in matrix form

$$dX(t) = AX(t)dt + H(t)dt + KdB(t), \tag{6}$$

where $dX = \begin{pmatrix} dX_1(t) \\ dX_2(t) \end{pmatrix}$ $A = \begin{pmatrix} 0 & 1 \\ \frac{-1}{CL} & \frac{-R}{L} \end{pmatrix}$ $H(t) = \begin{pmatrix} 0 \\ \frac{1}{L}V(t) \end{pmatrix}$ $K = \begin{pmatrix} 0 \\ \frac{\alpha}{L} \end{pmatrix}$,

and B_t is one dimensional Brownian motion. Rewrite equation (6) as

$$exp(-At)dX(t) = exp(-At)AX(t)dt + exp(-At)[H(t)dt + KdB_t], \qquad (7)$$

where for a general $n * n$ matrix A we define $exp(A) = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$. Applying a 2 dimensional version of the Ito formula for the function $g : [0, \infty) * R^2 \to R^2$ given by $g(t, x_1, x_2) = exp(-At) \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$ we obtain,

$$d(exp(-At)X(t)) = (-A)exp(-At)X(t)dt + exp(-At)dX(t) \qquad (8)$$

substituted in (7) this gives

$$exp(-At)X(t) - X(0) = \int_0^t exp(-As)H(s)ds + \int_0^t exp(-As)KdB_s \qquad (9)$$

or

$$X(t) = exp(At)[X(0) + exp(-At)KB_t + \int_0^t exp(-As)[H(s) + AKB_s]ds] \qquad (10)$$

by integration by parts [9].

Lemma 1: let $A = \begin{pmatrix} 0 & 1 \\ -a & -b \end{pmatrix}$ then

$$exp(At) = \frac{e^{-\lambda t}}{\xi}\{(\xi \cos(\xi t) + \lambda \sin(\xi t))I + A \sin(\xi t)\},$$

where,

$$\lambda = \frac{b}{2},\ \xi = \sqrt{a - \frac{b^2}{4}}.$$

Using this lemma the analytical solution can be achieved explicitly.
The solution X(t) is a random process and for it's expectation we have for every $t > 0$,

$$E(X(t)) = exp(At)E(X_0) + E(\int_0^t exp(A(t-s))H(s)ds + \int_0^t exp(A(t-s))KdB_s).$$

Since $E(\int_0^t f(s)dB_s) = 0$, then it can be easily shown that

$$E(X(t)) = exp(At)E(X_0) + \int_0^t e^{-A(s-t)}H(s)ds.$$

4. Numerical Solution and Simulation

To simulate $Q(t)$, numerical techniques have to be used. The simplest numerical scheme, the stochastic Euler scheme, is based on numerical methods for ordinary differential equations. The Euler scheme for system equation (5) is as follows [7],

$$\begin{cases} X_1(n+1) - X_1(n) = X_2(n) \cdot \Delta t_n \\ X_2(n+1) - X_2(n) = (\frac{-R}{L} X_2(n) - \frac{1}{CL} X_1(n) + \frac{1}{L} V(\Delta t_n))\Delta t_n + \frac{\alpha}{L} \Delta B_n. \end{cases} \quad (11)$$

Where $B_{n+1} - B_n \sim N(0, \Delta t_n),\ X_1(0) = Q(0),\ X_2(0) = I(0)$.

Example: Let us consider the RLC electrical circuit, when L,R,C and V(t) are constants, $I_0 = 0$. Using the Euler scheme we compute and graph the mean and three sample pass of stochastic solutions for $I(t)$ and $Q(t)$ with $\alpha = 10$ in fig. 1, and $\alpha = 100$ in fig. 2.

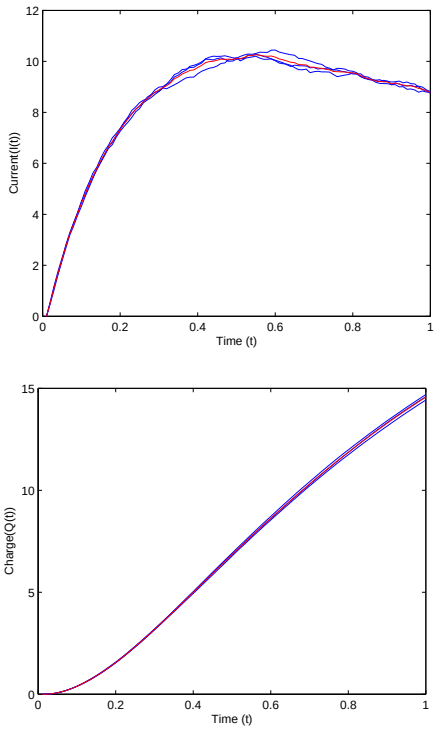

Fig. 1. The mean and three sample pass of stochastic solution with $\alpha = 10$.

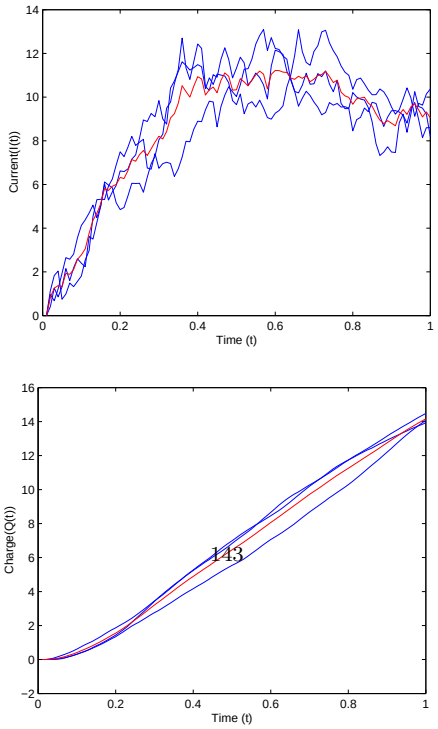

Fig. 2. The mean and three sample pass of stochastic solution with $\alpha = 100$.

References

[1] *L. Arnold, Stochastic Differential Equations, Theory and Applications, Johan Wiley Sons, (1974).*

[2] *K. Burrage and E. Platen, Rung-Kutta Methods for Stochastic Diferential Equations, Ann. Numer. Math.(1994), 1, 63-78.*

[3] *D. Ham, A. Hajimiri, complete noise analysis for CMOS switching mixtures via SDEs, IEEE Custom Integrated circuits conference, (2000), 439-442.*

[4] *Y. Kamarianakis, N. Frangos, Deterministic and Stochastic Diferential Equations modeling for electrical networks, Hellenic and European Research in Computational Mathematics Conference, Athens University of Economics Businness, (2001).*

[5] *W. Kampowsky, P. Rentrop and W. Schmidt, Classification and numerical simulation of electric circuits, Surveys Math. Indust, (1992), 23-65.*

[6] *C. Kim, E. K. Lee, Numerical Method for Solving Diferential Equations with Dichotomous noise, Physical Review, E73, 026101 (2006).*

[7] *P. E. Kloeden, E. Platen, Numerical solution of Stochastic Differential Equations, Springer, Berlin, (1995).*

[8] *E. Kolarova, Modeling RL electrical circuits by SDEs, International conference on computer as a tool, (2005), 1236-1238.*

[9] *B. Oksendal, Stochastic Diferential Equations, an introduction with applications, Springer-Verlage, (2000).*

[10] *C. Penski, A new numerical method for SDEs and its application in circuit simulation, Com. and Appl. Math, 115(2000), 461-470.*

Sum, Products and Ratios for Kibble's Bivariate Gamma Distribution

F. Fatahi, B. Mahaki and I. Almasi
Statistical Office of Country Ministry

University Department, Shahid Beheshti

University of Ilam

Kibble's bivariate gamma distribution has been around since the 1940s and has been applied in several areas of electrical and electronic engineering. However, it seems that no explicit expressions for its Sum, Products and Ratios are available. In this short note, we derive exact distributions of Sum, Products and Ratios and the corresponding moment properties are derived when X and Y Kibble's bivariate gamma distributions.
Keywords: Kibble's bivariate gamma distribution, Modified Bessel function, Modified Laguerre polynomial.

1. Introduction

Kibbles bivariate gamma distribution (Kibble, 1941) This distribution has received applications in several areas. Some of them are:
(a) Electric counter system: Lampard (1968) built up Kibbles distribution in the conditional manner, $h = f(x)g(y|x)$; his context was a system of two reversible counters (i.e. an input can either increase or decrease the cumulative count), with two Poisson inputs (an increase process and a decrease process). Output events occur when either of the cumulative counts decreases to zero. The sequence of time intervals between output events forms a Markov chain, and the joint distribution of successive intervals is of Kibbles form. Lampard also gave an interpretation of the same process in terms of a queueing system.
(b) Hydrology: Phatarford (1976) used Kibbles distribution as a model to describe summer and winter stream flows.
(c) Rain: As the gamma distribution is a popular univariate choice for the description of amount of rainfall, Izawa (1965) proposed Kibbles distribution to describe the joint distribution of rainfall at two nearby rain gauges.
(d) Wind gusts: Smith and Adelfang (1981) reported analysis of wind gust data using Kibbles distribution. The two variates were magnitude and length of the gust.

This distribution has given by the joint pdf

$$f_{X,Y}(x,y) = \frac{(xy)^{(\alpha-1)/2}}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}} exp\left(-\frac{x+y}{1-\rho}\right) I_{\alpha-1}\left(\frac{2\sqrt{xy\rho}}{1-\rho}\right) \qquad (1)$$

for $x > 0, y > 0, \alpha > 0$and $0 \leq \rho < 1$, where $I_v(.)$ denotes the modified Bessel function of the first kind of orderν .

Explicit expressions for the pdfs and moments of $Z = X + Y$,$R = X/Y$ and $U = XY$ for these distribution are derived in Sections 2-3 and Conditional PDF and moments are derived in Sections 4.The calculations involve several special functions, including the complementary error function defined by
the Jacobi polynomial defined by

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!}(1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^n}{dx^n}\{(1-x)^{-\alpha+n}(1+x)^{-\beta+n}\}$$

the modified Bessel function of the first kind defined by

$$I_n(x) = \sum_{k=0}^{\infty} \frac{x^{2k+\nu}}{2^{2k+\nu}\Gamma(\nu+k+1)k!}$$

for n is an integer.

$$I_\nu(x) = i^{-\nu}J_\nu(ix)\sum_{k=0}^{\infty} \frac{x^{2k+\nu}}{2^{2k+\nu}\Gamma(\nu+k+1)k!}$$

for ν is an real.
The modified Bessel function of the secund kind of order ν defined by

$$K_\nu(x) = \frac{\pi}{2}\frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\nu\pi)}$$

with $K_n(.)$ interpreted as the limit

$$K_n(x) = \lim_{\nu\to n} K_\nu(x)$$

the modified Laguerre polynomial defined by

$$L_n^\nu(x) = \frac{x^{-\nu}\exp(x)}{n!}\frac{d^n}{dx^n}\{x^{n+\nu}\exp(-x)\}$$

the Gauss hypergeometric function defined by

$${}_2F_1(a;b;c;x) = \sum_{k=0}^{\infty}\frac{(a)_k(b)_k}{(c)_k}\frac{x^k}{x!}$$

where $(\alpha)_k = \alpha(\alpha+1)...(\alpha+k-1)$denotes the ascending factorial. We also need the following important lemmas.

Lemma 1 :[Equation (2.15.3.2), Prudnikov et al., 1986, volume 2].For $\alpha+\nu>0$ and$p>c$,

$$\int_0^\infty x^{\alpha-1}\exp(-px)I_\nu(cx)dx$$
$$=p^{-(\alpha+\nu)}(\frac{c}{2})^\nu\frac{\Gamma(\nu+\alpha)}{\Gamma(\nu+1)}{}_2F_1(\frac{\alpha+\nu}{2},\frac{\alpha+\nu+1}{2};\nu+1;\frac{c^2}{p^2})$$

Lemma 2 :[Equation (2.15.2.6), Prudnikov et al., 1986, volume 2]. For $a>0$,$\beta>$ 0and $\nu>-1$,

$$\int_0^a x^{\nu+1}(a^2-x^2)^{\beta-1}I_\nu(cx)dx=2^{\beta-1}a^{\nu+\beta}c^{-\beta}\Gamma(\beta)I_{\nu+\beta}(ac)$$

Lemma 3: [Equation (2.15.5.4), Prudnikov et al., 1986, volume 2]. For $p>0$ and $\nu>-1$,

$$\int_0^\infty x^{\nu+2n+1}\exp(-px^2)I_\nu(cx)dx=\frac{n!c^\nu}{2^{\nu+1}p^{n+\nu+1}}\exp(\frac{c^2}{4p})L_n^\nu(-\frac{c^2}{4p})$$

Lemma 4: [Equation (2.263.1), Gradshteyn and Ryzhik, 2000]. For$m<2n$,

$$\int\frac{x^m}{(a+bx+cx^2)^{n+1/2}}dx=\frac{x^{m-1}}{(m-2n)c(a+bx+cx^2)^{n-1/2}}$$

$$-\frac{(2m-2n-1)b}{2(m-2n)c}\int\frac{x^{m-1}}{(a+bx+cx^2)^{n+1/2}}dx$$

$$-\frac{(m-1)a}{(m-2n)c}\int\frac{x^{m-2}}{(a+bx+cx^2)^{n+1/2}}dx$$

Lemma 5: [Equation (2.263.2), Gradshteyn and Ryzhik, 2000].

$$\int\frac{x^{2n}}{(a+bx+cx^2)^{n+1/2}}dx=-\frac{x^{2n-1}}{(2n-1)c(a+bx+cx^2)^{n-1/2}}$$

$$-\frac{b}{2c}\int\frac{x^{2n-1}}{(a+bx+cx^2)^{n+1/2}}dx+\frac{1}{c}\int\frac{x^{2n-2}}{(a+bx+cx^2)^{n+1/2}}dx$$

Lemma 6 :[Equation (2.263.4), Gradshteyn and Ryzhik, 2000]. For$n\geq 1$

$$\int\frac{1}{(a+bx+cx^2)^{n+1/2}}dx=\frac{2(2cx+b)}{(2n-1)(4ac-b^2)(a+bx+cx^2)^{n-1/2}}$$

$$\times\{1+\sum_{k=1}^{n-1}\frac{8^k(n-1)(n-2)...(n-k)c^k}{(2n-3)(2n-5)...(2n-2k-1)(4ac-b^2)^k}(a+bx+cx^2)^k\}$$

Lemma 7: [Equation (17.8.1.1),Alan Jeffrey , 2000].

$$I_{n+1/2}(x)=\frac{1}{\sqrt{2\pi x}}[e^x\sum_{k=0}^{n}\frac{(-1)^k(n+k)!}{k!(n-k)!(2x)^k}+e^{-x}(-1)^{n+1}\sum_{k=0}^{n}\frac{(n+k)!}{k!(n-k)!(2x)^k}]$$

Lemma 8:[Equation (17.10.2.3),Alan Jeffrey , 2000].

$$xI_{\nu-1}(x)=x\frac{d}{dx}[I_\nu(x)]+\nu I_\nu(x)$$

Lemma 9: [Equation (3.471.9), Alan Jeffrey,Daniel Zwillinger 2000].

$$\int_0^\infty x^{\nu-1}e^{-\frac{\beta}{x}-\gamma x}dx=2(\beta/\gamma)^{\nu/2}K_\nu(2\sqrt{\beta\gamma})$$

for$\beta>0$ and $\gamma>0$.

2. Pdfs

Theorems(1),(2) and (3) derive the pdfs of $Z=X+Y$, $R=X/Y$ and $U=XY$ when X and Y are distributed according to Eq.(1).

Theorem 1. If X and Y are jointly distributed according to Eq.(1) then

$$f_Z(z)=\frac{\sqrt{\pi}2^{1/2-\alpha}z^{\alpha-1/2}\rho^{1/4-\alpha/2}}{\Gamma(\alpha)\sqrt{1-\rho}}\exp(-\frac{z}{1-\rho})I_{\alpha-1/2}\left(\frac{z\sqrt{\rho}}{1-\rho}\right) \qquad (2)$$

for $0<z<\infty$.

Proof.From Eq.(1), the pdf of $Z=X+Y$ becomes

$$f(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx=\int_0^z f(x,z-x)dx$$

$$=\frac{exp(-z/1-\rho)}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}}\int_0^z(x(z-x))^{(\alpha-1)/2}I_{\alpha-1}\left(\frac{2\sqrt{x(z-x)\rho}}{1-\rho}\right)dx$$

Thus the pdf of Z can be written as

$$f(z)=\frac{exp(-z/1-\rho)}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}}J(z) \qquad (3)$$

where

$$J(z)=\int_0^z(x(z-x))^{(\alpha-1)/2}I_{\alpha-1}\left(\frac{2\sqrt{x(z-x)\rho}}{1-\rho}\right)dx$$

Substituting $u=\sqrt{x(z-x)}$, the integral J (z) can be rewritten as

$$J(z)=\int_0^{z/2}2u^\alpha(z^2-4u^2)^{-1/2}I_{\alpha-1}\left(\frac{2\sqrt{\rho}}{1-\rho}u\right)du \qquad (4)$$

Direct application of Lemma(2) shows that Eq.(4),and Substituting $v = 2u$ can be calculated as

$$J(z) = \int_0^z 2^{-\alpha} v^{\alpha} (z^2 - v^2)^{-1/2} I_{\alpha-1}\left(\frac{\sqrt{\rho}}{1-\rho} v\right) du \tag{5}$$

The result of the theorem follows by combining Eq.(3) and Eq.(5).□

Theorem 2. If X and Y are jointly distributed according to Eq.(1) then

$$f_R(r) = \frac{\Gamma(2\alpha)(1-\rho)^{\alpha} r^{\alpha-1}(1+r)}{\Gamma^2(\alpha)((1+r)^2 - 4\rho r)^{\alpha+1/2}} \tag{6}$$

for $0 < R < \infty$.

Proof.Using Eq.(1), one can write

$$f(r,y) = \frac{r^{(\alpha-1)/2} y^{\alpha}}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}} \exp\left(-\frac{y(r+1)}{1-\rho}\right) I_{\alpha-1}\left(\frac{2\sqrt{r\rho}}{1-\rho} y\right) \tag{7}$$

Thus

$$f(r) = \frac{r^{(\alpha-1)/2}}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}} \int_0^{\infty} y^{\alpha} \exp\left(-\frac{y(r+1)}{1-\rho}\right) I_{\alpha-1}\left(\frac{2\sqrt{r\rho}}{1-\rho} y\right)$$

Direct application of Lemma(1) shows that one can calculate f(r) as

$$f(r) = \frac{r^{(\alpha-1)/2}}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}} \left(\frac{r+1}{1-\rho}\right)^{-2\alpha} \left(\frac{\sqrt{r\rho}}{1-\rho}\right)^{\alpha-1} \frac{\Gamma(2\alpha)}{\Gamma(\alpha)} {}_2F_1(\alpha, \alpha+1/2, \alpha; \frac{4\rho r}{(1+r)^2}) \tag{8}$$

Upon using the property that

$${}_2F_1(a,b,a;x) = (1-x)^{-b}$$

one can reduce (8) to The result of the theorem.□

Using special properties of the Bessel function of the first kind, one can derive elementary forms for the pdfs in Eq.(2) . This is illustrated in the corollary below.

Lemma 10. If X and Y are jointly distributed according to Eq.(1) and if $\alpha \geq 1$ is an integer then

$$f_Z(z) = \frac{z^{\alpha-2}(1-\rho)^{1/2}}{\sqrt{\pi}\Gamma(\alpha) 2^{\alpha} \rho^{(\alpha+1)/2}} \exp(-\frac{z}{1-\rho}) \Big[\exp(\frac{z\sqrt{\rho}}{1-\rho}) \sum_{k=0}^{n} \frac{(-1)^k (n+k)!}{k!(n-k)!} \frac{(1-\rho)^k \rho^{-k/2}}{(2z)^k} (n + z\frac{\sqrt{\rho}}{1-\rho}) + (-1)^{n+1} \exp(-\frac{z\sqrt{\rho}}{1-\rho}) \sum_{k=0}^{n} \frac{(n+k)!}{k!(n-k)!} \frac{(1-\rho)^k \rho^{-k/2}}{(2z)^k} (n - z\frac{\sqrt{\rho}}{1-\rho}) \Big] \tag{9}$$

Proof.It follows from Lemma(7) Then according Lemma(8) that

$$xI_{n-1/2}(x) = x\frac{d}{dx}[I_{n+1/2}(x)] + (n+1/2)I_{n+1/2}(x)$$

Then

$$xI_{n-1/2}(x) = \frac{1}{\sqrt{2\pi x}}[e^x \sum_{k=0}^{n} \frac{(-1)^k(n+k)!}{k!(n-k)!(2x)^k}(n+x-k)$$

$$+e^{-x}(-1)^{n+1}\sum_{k=0}^{n}\frac{(n+k)!}{k!(n-k)!(2x)^k}(n-x-k)] \quad (10)$$

The result of the lemma follows by combining Eq.(2) and Eq.(10).□

Theorem 3. If X and Y are jointly distributed according to Eq.(1) then

$$f_U(u) = \frac{u^{(\alpha-1)/2}}{\Gamma(\alpha)(1-\rho^{(\alpha-1)/2})} I_{\alpha-1}\left(\frac{2\sqrt{u\rho}}{1-\rho}\right) K_0(\frac{2\sqrt{u}}{1-\rho}) \quad (11)$$

for $0 < u < \infty$.

Proof.From Eq.(1), the joint pdf of (X, U) becomes

$$f(x,u) = \frac{u^{(\alpha-1)/2}}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}}\frac{1}{x}\exp\left(-\frac{x}{1-\rho}-\frac{u}{(1-\rho)x}\right) I_{\alpha-1}\left(\frac{2\sqrt{u\rho}}{1-\rho}\right)$$

Thus, the pdf of U can be written as

$$f(U) = \frac{u^{(\alpha-1)/2}}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}} I_{\alpha-1}\left(\frac{2\sqrt{u\rho}}{1-\rho}\right)$$
$$\times \int_0^\infty \frac{1}{x}\exp\left(-\frac{x}{1-\rho}-\frac{u}{(1-\rho)}.\frac{1}{x}\right)dx \quad (12)$$

Direct application of Lemma(9) shows that Eq.(12) can be calculated as

$$\int_0^\infty \frac{1}{x}\exp\left(-\frac{x}{1-\rho}-\frac{u}{(1-\rho)}.\frac{1}{x}\right)dx = K_0(\frac{2\sqrt{u}}{1-\rho})$$

Thus, the pdf of U can be written as

$$f_U(u) = \frac{u^{(\alpha-1)/2}}{\Gamma(\alpha)(1-\rho^{(\alpha-1)/2})} I_{\alpha-1}\left(\frac{2\sqrt{u\rho}}{1-\rho}\right) K_0\left(\frac{2\sqrt{u}}{1-\rho}\right) \quad (13)$$

where

$$K_0(x) = -[\ln\frac{x}{2}+\gamma]I_0(x)+\frac{1/4x^2}{1!}+(1+1/2)\frac{(1/4x^2)^2}{(2!)^2}+(1+1/2+1/3)\frac{(1/4x^2)^3}{(3!)^2}+...$$

where γ is Eulers constant (also known as the EulerMascheroni constant) defined as

$$\gamma = \lim_{n\to\infty}\left(1+\frac{1}{2}+\frac{1}{3}+...+\frac{1}{n}-\ln n\right) = 0.57721566...$$

the proof is complement.□

3. Moments

Now, we derive the moments of $Z = X + Y$, $U = XY$ and $R = X/Y$ when X and Y are distributed according to Eq.(1). We need the following theorem.

Theorem 4. If X and Y are jointly distributed according to Eq.(1) then [6],

$$E(X^m Y^n) = \frac{n!\Gamma(m+\alpha)(1-\rho)^n}{\Gamma(\alpha)} P_n^{(\alpha-1,m-n)}(\frac{1+\rho}{1-\rho}) \tag{14}$$

for $m \geq 1 \text{and} n \geq 1$.

Corollary1. If X and Y are jointly distributed according to Eq.(1)., then

$$Cov(X,Y) = \alpha\rho \quad , \quad Corr(x,y) = \rho \tag{15}$$

The moments of $Z = X + Y$ are now simple consequences of this lemma as illustrated in Theorem(5). The moments of $R = X/Y$ require a separate treatment as shown by Theorem(4).

Theorem 5. If X and Y are jointly distributed according to Eq.(1) then

$$E(Z^n) = \sum_{k=0}^{n} \frac{n!\Gamma(n-k+\alpha)(1-\rho)^k}{(n-k)!\Gamma(\alpha)} P_n^{(\alpha-1,n-2k)}(\frac{1+\rho}{1-\rho}) \tag{16}$$

for$n \geq 1$.

Proof. The result in Eq.(16) follows by writing

$$E((X+Y)^n) = \sum_{k=0}^{n} \binom{n}{k} E(X^{n-k}Y^k)$$

and applying theorem(4) to each expectation in the sum.□

Theorem 6. If X and Y are jointly distributed according to Eq.(1) and if α is an integer then

$$E(R^n) = \frac{\Gamma(2\alpha)(1-\rho)^\alpha}{\Gamma^2(\alpha)} \sum_{k=0}^{n+\alpha-1} (-1)^k \binom{n+\alpha-1}{k} J(n++\alpha-k) \tag{17}$$

for$n \geq 1$.where J (m) satisfies the recurrence relations for $n - k < \alpha$

$$J(n+\alpha-k) = -\frac{1}{n-\alpha-k} + \frac{(2n-2k-1)2\rho}{n-\alpha-k} J(n+\alpha-k-1)$$

$$-\frac{(n+\alpha-k-1)4\rho}{n-\alpha-k} J(n+\alpha-k-2) \tag{18}$$

and for $n - k = \alpha$

$$J(2\alpha) = \frac{1}{2\alpha-1} + 2\rho J(2\alpha-1) + J(2\alpha-2) \tag{19}$$

for$m \neq 2\alpha$,with the initial values

$$J(0) = \frac{2\rho - 1}{(2\alpha - 1)4\rho(1-\rho)}\{1 + \sum_{k=1}^{\alpha-1} \frac{(\alpha-1)(\alpha-2)...(\alpha-k)}{(2\alpha-3)(2\alpha-5)...(2\alpha-2k-1)2^k\rho^k(1-\rho)^k}\} \tag{20}$$

and $J(1) = \frac{1}{2\alpha-1} + 2\rho J(0)$.

Proof. It follows from Eq.(6) that

$$E(R^n) = \frac{\Gamma(2\alpha)(1-\rho)^\alpha}{\Gamma^2(\alpha)} \int_0^\infty \frac{r^{n+\alpha-1}(1+r)}{((r+1)^2 - 4\rho r)^{\alpha+1/2}} dr$$

which follows after setting$1 + r = w$ then

$$\frac{\Gamma(2\alpha)(1-\rho)^\alpha}{\Gamma^2(\alpha)} \int_1^\infty \frac{(w-1)^{n+\alpha-1}w}{(w^2 - 4\rho w + 4\rho)^{\alpha+1/2}} dw$$

$$= \frac{\Gamma(2\alpha)(1-\rho)^\alpha}{\Gamma^2(\alpha)}$$

$$\times \int_1^\infty \frac{w}{(w^2 - 4\rho w + 4\rho)^{\alpha+1/2}} \left\{ \sum_{k=0}^{n+\alpha-1} (-1)^k \binom{n+\alpha-1}{k} w^{n+\alpha-k-1} \right\} dw$$

$$= \frac{\Gamma(2\alpha)(1-\rho)^\alpha}{\Gamma^2(\alpha)} \sum_{k=0}^{n+\alpha-1} (-1)^k \binom{n+\alpha-1}{k} J(n+\alpha-k)$$

where

$$J(m) = \int_1^\infty \frac{w^m}{(4\rho - 4\rho w + w^2)^{\alpha+1/2}} dw$$

. This establishes the result in Eq.(17). The recurrence relations in Eq.(18) and Eq.(20) follow by applying Lemmas(4) and (5). The initial value in Eq.(20) follows by applying Lemma(6). Finally, since it follows that $J(1) = \frac{1}{2\alpha-1} + 2\rho J(0)$.□

Theorem7. If U distributed according to Eq.(11) then

$$E(U^n) = \frac{n!\Gamma(n+\alpha)(1-\rho)^n}{\Gamma(\alpha)} P_n^{(\alpha-1,0)}(\frac{1+\rho}{1-\rho}) \tag{21}$$

for $n \geq 1$.

Proof. The result in Eq.(14) follows by writing $m = n$.□

4. Conditional PDF and moment's

Theorems(8) and(9) derive the conditional distributions corresponding to kibble's bivariate gamma distribution

Theorem 8.If X and Y are jointly distributed according to Eq.(1), the conditional

pdf of X given $Y = y$ is given by

$$f(x|y) = \frac{1}{(1-\rho)\rho^{(\alpha-1)/2}} \left(\frac{x}{y}\right)^{(\alpha-1)/2} \exp\left(-\frac{1}{1-\rho}(x+\rho y)\right) I_{\alpha-1}\left(\frac{2\sqrt{xy\rho}}{1-\rho}\right) \tag{22}$$

Proof.

$$f_X(x) = \frac{(x)^{(\alpha-1)/2} e^{-x/(1-\rho)}}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}} \int_0^\infty y^{(\alpha-1)/2} exp(-y/(1-\rho)) I_{\alpha-1}\left(\frac{2\sqrt{xy\rho}}{1-\rho}\right)$$

Substituting $u = \sqrt{y}$ and then using the lemma (3) , one can obtain the expression

$$f(x) = \frac{x^{(\alpha-1)/2} e^{-x/(1-\rho)}}{\Gamma(\alpha)(1-\rho)\rho^{(\alpha-1)/2}} J(x) \tag{23}$$

where

$$J(x) = 2\int_0^\infty u^\alpha exp\left(-\frac{u^2}{1-\rho}\right) I_{\alpha-1}\left(\frac{2\sqrt{x\rho}}{1-\rho}u\right) du$$

Direct application of Lemma(3) shows that $J(x)$,can be calculated as

$$J(x) = (1-\rho)(x\rho)^{(\alpha-1)/2} exp\left(\frac{\rho}{1-\rho}x\right) \tag{24}$$

By combining Eq.(23) and Eq.(24).

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x) \tag{25}$$

for $0 \leq x < \infty$.

Corollary 3. In a similar manner

$$f(y) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} \exp(-y) \tag{26}$$

for $0 \leq y < \infty$.

□

Theorem 9.For the conditional pdf of X given Y =y is given by the Eq.(23),

$$E\left(X^m|Y=y\right) = m!(1-\rho)^m L_m^{(\alpha-1)}\left(-\frac{\rho}{1-\rho}y\right) \tag{27}$$

Proof.

$$E\left(X^m|Y=y\right) = \frac{e^{-\frac{\rho}{1-\rho}y}}{y^{(\alpha-1)/2}(1-\rho)\rho^{(\alpha-1)/2}} \int_0^\infty x^{m+\frac{\alpha-1}{2}} e^{-\frac{x}{1-\rho}} I_{\alpha-1}\left(\frac{2\sqrt{xy\rho}}{1-\rho}\right)$$

Substituting $u = \sqrt{x}$ and then using the lemma (3) , one can obtain the expression

$$E\left(X^m|Y=y\right) = \frac{e^{-\frac{\rho}{1-\rho}y}}{y^{(\alpha-1)/2}(1-\rho)\rho^{(\alpha-1)/2}} 2\int_0^\infty u^{2m+\alpha} e^{-\frac{u^2}{1-\rho}} I_{\alpha-1}\left(\frac{2\sqrt{y\rho}}{1-\rho}u\right)$$

$$= m!(1-\rho)^m L_m^{(\alpha-1)}\left(-\frac{\rho}{1-\rho}y\right)$$

proof is complement.□

References

1. I.S.Gradshteyn and I.M.Ryzhik, *"Table of Integrals,Series and Products"*, sixth edition, San Diego:Academic Press,(2000).
2. T.Izawa, *"Two or multi-dimensional gamma-type distribution and its application to rainfall data"*, Vol.Papers in Meteorology and Geophysics, 15 (1965), 167200.
3. Alan Jeffrey., *"HAND BOOK OF MATHEMATICAL FORMULAS AND INTEGRALS"*, (Second Edition)ACADEMIC PRESS(2000).
4. Alan Jeffrey and Daniel Zwillinger, *"Table of Integrals, Series, and Products"*, (6 Edition)Academic Press(1996).
5. D.G. Lampard., *"A stochastic process whose successive intervals between events form a first order Markov chain1"*, Journal of Applied Probability, 5 (1968), 648668.
6. S.Nadarajah and S.Kotz , *"Product moments of kibble's bivariate gamma distribution"*, Circuits Systems Singnal Processing Vol.25,NO.4,Pp.467-570 ,(2006).
7. A.P.Prudnikov,Y.A.Brychkov and O.I.Marichev., *"Integrals and Series (volumes 1,2 and 3)"*, Amsterdam:Gordon and Breach Science Publishers,(1986).
8. O.E. Smith and S.I.Adelfang, *"Gust model based on the bivariate gamma probability distribution"*Journal of Spacecraft and Rockets, 18 (1981), 545549.

Asymptotic Deficiency for Inhomogeneous Poisson Processes

Kh. Fazli (Invited)

Department of Statistics, University of Kurdistan

In this work, based on a realization of an inhomogeneous Poisson process whose intensity function depends on an unknown real parameter, we consider a simple null hypothesis against a sequence of close (contiguous) one-sided alternatives. The main object is to obtain the asymptotic deficiency of the score test with respect to the Neyman-Pearson test.

Keywords: Inhomogeneous Poisson processes, Hypotheses testing, asymptotic deficiency, Pitman's asymptotic relative efficiency, Power loss, Score test.

1. Introduction and Preliminaries

Inhomogeneous Poisson process is widely used in many fields. The diverse choice of intensity function makes it a suitable model for many real phenomena. Estimation theory for spatial Poisson process was extensively developed (see [3], [9] and the references therein). For the problem of goodness of fit testing with the simple basic hypothesis and a nonparametric alternative see [8]. One can refer to [4] and [2] where a simple Poissonian null hypothesis is tested against a large classes of alternatives of the type of stationary point processes. In the present work we will consider a simple null hypothesis against the one sided parametric alternative when the parameter is one dimensional. The second order efficiency of the *Rao score* test has been proved in [5]. The *power loss* of this test which measures the asymptotic gap of its power with the Neyman-Pearson test under the close alternatives is obtained in [7]. In this work we will obtain the *asymptotic deficiency* of score test with respect to the Neyman-Pearson test, *i.e.*, the asymptotic "number" of additional observations needed for the score test to have the same power as the most powerful test under the close alternatives.

Let $X^{(n)}$ be an inhomogeneous Poisson process observed on some increasing subsets $\mathbb{A}_n,\ n = 1, 2, \ldots$ of d dimensional Euclidian space $\mathbb{R}^d$ with intensity function $S(\vartheta, x), x \in \mathbb{A}_n$ depending on the real parameter $\vartheta \in \Theta$. Let $\left\{\boldsymbol{P}_{\vartheta}^{(n)},\ \vartheta \in \Theta\right\}$ denote the parametric family of distributions of the random element $X^{(n)}$. Based on $X^{(n)}$ we are interested in the problem of testing $\mathcal{H}_0 :\ \vartheta = \vartheta_0$ against $\mathcal{H}_1 :\ \vartheta > \vartheta_0$,

where ϑ_0 is a given value in the parameter space Θ. Let us fix a preassigned level of significance α between 0 and 1 and consider the class of tests ϕ_n at asymptotic level α, *i.e.*, $\lim_{n\to\infty} \mathbf{E}_{\vartheta_0}\phi_n\left(X^{(n)}\right) = \alpha$, where $\mathbf{E}_\vartheta$ denotes the mathematical expectation with respect to the probability measure $\boldsymbol{P}_\vartheta^{(n)}$. Following the Pitman's approach ([10]) we consider the so-called close (contiguous) alternatives, *i.e.*, a sequence of simple alternatives $\mathcal{H}_u : \vartheta = \vartheta_0 + \varphi_n\, u$, where $u > 0$ and the normalizing factor φ_n converges to zero with a certain rate. The most powerful test for $\mathcal{H}_0 : \vartheta = \vartheta_0$ against $\mathcal{H}_u : \vartheta = \vartheta_0 + \varphi_n u$ is the Neyman-Pearson test, denoted by $\tilde{\phi}_n$, based on the log-likelihood ratio. By using the Edgeworth type expansion we can construct a *second order efficient* test, *i.e.*, a test ϕ_n^* such that for any $K > 0$,

$$\sup_{0\le u\le K} |\mathbf{E}_{\vartheta_u}\tilde{\phi}_n\left(X^{(n)}\right) - \mathbf{E}_{\vartheta_u}\phi_n^*\left(X^{(n)}\right)| = O(\varepsilon_n^2),$$

for some sequence $\varepsilon_n \to 0$, where $\vartheta_u = \vartheta_0 + \varphi_n\, u$ ([5], see also (3)). Indeed, ϕ_n^* is the score test based on the derivative of the log-likelihood ratio with respect to ϑ at $\vartheta = \vartheta_0$ (see (2) and (4)). Hence to measure the performance ϕ_n^* it is natural to consider the *power loss* of ϕ_n^* with respect to the most powerful test $\tilde{\phi}_n$, which is defined by

$$r(u) = \lim_{n\to\infty} \varepsilon_n^{-2}\left(\mathbf{E}_{\vartheta_u}\tilde{\phi}_n\left(X^{(n)}\right) - \mathbf{E}_{\vartheta_u}\phi_n^*\left(X^{(n)}\right)\right), \tag{1}$$

for $u > 0$, see [7]. The power loss is closely related to the *deficiency* of ϕ_n^* with respect to $\tilde{\phi}_n$ which is the "number" of additional observations, denoted by d_n, needed for this test to have the same power as the most powerful test. This requires to take into account higher order terms in the Edgeworth expansions of the distribution functions of the test statistics. The main object of this work is to obtain the explicit form of asymptotic deficiency of the score test ϕ_n^*, *i.e.*, to obtain $d = \lim_{n\to\infty} d_n$, based on a realization $X^{(n)}$ of a nonhomogeneous Poisson process with intensity function $S(\vartheta, x), x \in \mathbb{A}_n$. We apply these results in several models. For the deficiency and power loss results in the *i.i.d.* case see [1] and the references therein. The definitions of the spatial Poisson processes as well as their properties and examples can be found in many books devoted to point processes (see, e.g., Daley and Vere-Jones [3]). Let $\boldsymbol{P}_\vartheta^{(n)}$ denote the probability law induced by the random element (realization) $X^{(n)}$ of a Poisson process observed on $\mathbb{A}_n$ with intensity function $S(\vartheta, x),\ x \in \mathbb{A}_n$. If the intensity measures $\Lambda_{\vartheta_0}^{(n)}, \Lambda_\vartheta^{(n)}$ defined by $\Lambda_\vartheta^{(n)}(\mathbb{B}) = \int_{\mathbb{B}} S(\vartheta, x)\ \mathrm{d}x$, where $\mathbb{B} \subseteq \mathbb{A}_n$, are equivalent then the corresponding probability measures $\boldsymbol{P}_{\vartheta_0}^{(n)}$ and $\boldsymbol{P}_\vartheta^{(n)}$ are equivalent with the following log-likelihood ratio

$$\ln \frac{\mathrm{d}\boldsymbol{P}_\vartheta^{(n)}}{\mathrm{d}\boldsymbol{P}_{\vartheta_0}^{(n)}}\left(X^{(n)}\right) = \int_{\mathbb{A}_n} \ln \frac{S(\vartheta, x)}{S(\vartheta_0, x)}\ X^{(n)}(\mathrm{d}x) - \int_{\mathbb{A}_n} [S(\vartheta, x) - S(\vartheta_0, x)]\ \mathrm{d}x \tag{2}$$

See [9] page 28. In the above relation the stochastic integral with respect to $X^{(n)}$ is defined by $\int_{\mathbb{A}_n} f(x)\ X^{(n)}(\mathrm{d}x) = \sum_{x_i \in \mathbb{A}_n} f(x_i)$, for a large class of functions f defined on $\mathbb{A}_n$ where $\{x_i\}$ are the *events* (random points) of the Poisson process ([9]). Following [9], we present the conditions under which the distribution function $F_n(y) = \boldsymbol{P}_\vartheta^{(n)} \{I(f_n) < y\}$ of the stochastic integral $I(f_n) = \int_{\mathbb{A}_n} f_n(x)\ \pi^{(n)}(\mathrm{d}x)$, admits a three terms Edgeworth type expansion, where $\pi^{(n)}(\mathrm{d}x) = X^{(n)}(\mathrm{d}x) - S(\vartheta, x)\,\mathrm{d}x$, is the centered Poisson process. The cumulant $\gamma_{r,n}$ of order r of $I(f_n)$ is given by $\gamma_{r,n} = \int_{\mathbb{A}_n} f_n(x)^r\ S(\vartheta, x)\,\mathrm{d}x$, $r = 2, 3, 4, ...$ See [9], page 20. Without loss of generality we suppose the variance of $I(f_n)$ is equal to 1, *i.e.*, $\gamma_{2,n} = 1$. The expansion is obtained under the following two conditions:

- $\mathcal{B}_1$. *There exists a sequence of real numbers* $\varepsilon_n \to 0$, *as* $n \to \infty$ *and constants* $C_r > 0, r = 3, 4, 5$, *such that* $\int_{\mathbb{A}_n} |f_n(x)|^r\ S(\vartheta, x)\,\mathrm{d}x \le C_r\ \varepsilon_n^{r-2}$.
- $\mathcal{B}_2$. *There exist constants* $\gamma \ge 5/2$ *and* $c_0 > 0$ *enough small such that* $\inf_{\frac{c_0 \varepsilon_n^{-1}}{2} < t < \frac{\varepsilon_n^{-2}}{2}} \int_{\mathbb{A}_n} \sin^2(t f_n(x))\ S(\vartheta, x)\,\mathrm{d}x \ge \gamma \ln \varepsilon_n^{-1}$ *for all large* n.

Note that $\gamma_{r,n} = O(\varepsilon_n^{r-2}),\ r = 3, 4$ by $\mathcal{B}_1$. The condition $\mathcal{B}_1$ concerns the cumulants of the stochastic integral $I(f_n)$ and $\mathcal{B}_2$ is Cramér-type condition which implies that the tails of the distribution of $I(f_n)$ are enough small. Let us introduce the Hermite polynomials: $H_2(y) = y^2 - 1$, $H_3(y) = y^3 - 3y$, $H_5(y) = y^5 - 10y^3 + 15y$.

Theorem 1.1. *Let the conditions* $\mathcal{B}_1$, $\mathcal{B}_2$ *be fulfilled. Then (i) uniformly in* $y \in \mathbb{R}$

$$F_n(y) = \mathcal{N}(y) - \frac{\gamma_{3,n}}{3!} H_2(y)\,\mathrm{n}(y) - \frac{\gamma_{4,n}}{4!} H_3(y)\,\mathrm{n}(y) - \frac{\gamma_{3,n}^2}{72} H_5(y)\,\mathrm{n}(y) + O(\varepsilon_n^3),$$

for all n *large. Here* $\mathcal{N}(y)$ *and* $\mathrm{n}(y)$ *denote the distribution and density functions of the standard Gaussian law, respectively; (ii) for given* $0 < \alpha < 1$ *the equation* $F_n(y) = 1 - \alpha + O(\varepsilon_n^3)$ *has a solution* $y = c_{n,\alpha} = z_\alpha + \frac{\gamma_{3,n}}{3!} H_2(z_\alpha) + \frac{\gamma_{4,n}}{4!} H_3(z_\alpha) + \frac{\gamma_{3,n}^2}{72} H_5(z_\alpha)$, *where* z_α *is the* $1 - \alpha$ *quantile of standard Gaussian law, i.e.,* $\boldsymbol{P}\{\zeta > z_\alpha\} = \alpha$ *where* $\zeta \sim N(0, 1)$.

For the proof see [**?**], page 40. The first part of the theorem is a special case of a general theorem given by Kutoyants, where the expansion is obtained by the powers of ε_n up to order ε_n^k, $k = 1, 2, ...$ ([9], page 131).

2. Pitman's Asymptotic Relative Efficiency

Let $\beta_n(\vartheta_u, \phi_n)$ denote the power of a test ϕ_n at the local alternative $\vartheta_u = \vartheta_0 + \varphi_n u$, *i.e.*, $\beta_n(\vartheta_u, \phi_n) = \mathbf{E}_{\vartheta_u} \phi_n(X^{(n)})$. We introduce the score test ϕ_n^*

$$\phi_n^*\left(X^{(n)}\right) = \begin{cases} 1, & \text{if } \Delta_n(\vartheta_0) > c_n \\ 0, & \text{if } \Delta_n(\vartheta_0) \le c_n, \end{cases} \tag{3}$$

based on the score statistic

$$\Delta_n(\vartheta_0) = \varphi_n \int_{\mathbb{A}_n} \frac{S^{(1)}(\vartheta_0, x)}{S(\vartheta_0, x)} \pi^{(n)}(\mathrm{d}x), \tag{4}$$

for some threshold c_n (see the next section) where $\pi^{(n)}(\mathrm{d}x) = X^{(n)}(\mathrm{d}x) - S(\vartheta_0, x)\,\mathrm{d}x$ is the centered Poisson process, $S^{(1)}(\vartheta, x)$ denotes the first derivative of $S(\vartheta, x)$ with respect to ϑ, and $\varphi_n = I_n(\vartheta_0)^{-1/2}$, where $I_n(\vartheta_0)$ is the *Fisher information* at ϑ_0

$$I_n(\vartheta_0) = \int_{\mathbb{A}_n} \frac{S^{(1)}(\vartheta_0, x)^2}{S(\vartheta_0, x)}\,\mathrm{d}x.$$

We suppose that $\varphi_n \to 0$ as $n \to \infty$. By the Neyman-Pearson lemma the most powerful test for $\mathcal{H}_0 : \vartheta = \vartheta_0$ against the local alternative $\mathcal{H}_u : \vartheta = \vartheta_0 + \varphi_n u$ with $u > 0$, is given by

$$\tilde{\phi}_n\left(X^{(n)}\right) = \begin{cases} 1, & \text{if } \Lambda_n(u) > b_n(u) \\ 0, & \text{if } \Lambda_n(u) < b_n(u) \end{cases}$$

where $\Lambda_n(u) = \ln \frac{\mathrm{d}\boldsymbol{P}^{(n)}_{\vartheta_u}}{\mathrm{d}\boldsymbol{P}^{(n)}_{\vartheta_0}}\left(X^{(n)}\right)$ is the log-likelihood ratio and the constant $b_n(u)$ can be chosen such that $\mathbf{E}_{\vartheta_0}\tilde{\phi}_n\left(X^{(n)}\right) = \alpha$. The power of $\tilde{\phi}_n$ as a function of u is called the *envelope power function* . Note that $\tilde{\phi}_n$ is not a test for the main hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ because it depends on the parameter u. The *Pitman's asymptotic relative efficiency* of ϕ^*_n with respect to $\tilde{\phi}_n$, is defined as the limit $e = \lim_{n\to\infty} \frac{k_n}{n}$, where k_n denotes the "number" of observations needed for the score test to have the same power under the local alternative ϑ_u as the most powerful test. As the Pitman's asymptotic efficiency problem is in the accuracy level $o(1)$, hence instead of the equation $\beta_{k_n}\left(\vartheta_u, \phi^*_{k_n}\right) = \beta_n(\vartheta_u, \tilde{\phi}_n)$, it is sufficient to consider $\beta_{k_n}\left(\vartheta_u, \phi^*_{k_n}\right) = \beta_n(\vartheta_u, \tilde{\phi}_n) + o(1)$, as $n \to \infty$. It is well known that, under *local asymptotic normality* (LAN) conditions at ϑ_0, the Rao score test is *locally asymptotically uniformly most powerful*, *i.e.*, for any $\sup_{0\le u\le K}|\beta_n(\vartheta_u, \tilde{\phi}_n) - \beta_n(\vartheta_u, \phi^*_n)| = o(1)$ for any $K > 0$, $u > 0$ and the powers admit the representations

$$\beta_n(\vartheta_u, \phi^*_n) = \mathcal{N}(u - z_\alpha) + o(1), \quad \beta_n(\vartheta_u, \tilde{\phi}_n) = \mathcal{N}(u - z_\alpha) + o(1).$$

Now we can write

$$\begin{aligned}\beta_{k_n}\left(\vartheta_u, \phi^*_{k_n}\right) =& \beta_{k_n}\left(\vartheta_0 + \varphi_n u, \phi^*_{k_n}\right) = \beta_{k_n}(\vartheta_0 + \varphi_{k_n}\frac{\varphi_n}{\varphi_{k_n}}u, \phi^*_{k_n}) = \\ &= \beta_{k_n}\left(\vartheta_0 + \varphi_{k_n} u_n, \phi^*_{k_n}\right) = \beta_{k_n}\left(\vartheta_{u_n}, \phi^*_n\right),\end{aligned}$$

where $u_n = \frac{\varphi_n}{\varphi_{k_n}} u$. Setting $k_n = n + o(n)$ and assume $\varphi_n \sim C\, n^{-p}$ for some constants $C > 0$ and $p > 0$, we have $\lim_{n\to\infty} u_n = u$. By considering the above representations of the powers, we get

$$\beta_{k_n}\left(\vartheta_u, \phi^*_{k_n}\right) = \beta_{k_n}\left(\vartheta_0 + \varphi_{k_n}\, u_n, \phi^*_{k_n}\right) = \mathcal{N}(u_n - z_\alpha) + o(1) =$$
$$= \mathcal{N}(u - z_\alpha) + o(1) = \beta_n\left(\vartheta_u, \tilde{\phi}_n\right) + o(1),$$

as desired. Hence the Pitman's asymptotic efficiency $e = 1$. This quantity doesn't say anything about the asymptotic of the differences $d_n = k_n^q - n^q$ for some $q > 0$. For $q = 1$ the quantity $d_n = k_n - d_n$ is the so-called *deficiency* of ϕ^*_n with respect to the Neyman-Pearson test $\tilde{\phi}_n$, *i.e.*, the "number" of additional observations needed for the score test to have the same power as the most powerful test (see [1]). Our goal is to obtain the *asymptotic deficiency* of ϕ^*_n defined by

$$d = \lim_{n\to\infty} d_n.$$

3. Asymptotic Deficiency

Let $S^{(j)}(\vartheta, x)$ denote the $j-$th derivative of $S(\vartheta, x)$ with respect to ϑ. We introduce the following conditions:

$\mathcal{D}_1$. The intensity function $S(\vartheta, x)$ is two times differentiable with respect to ϑ in a right neighborhood of ϑ_0 and the normalizing factor $\varphi_n \sim C\, n^{-p}$ for some $p > 0$ and $C > 0$.

$\mathcal{D}_2$. The conditions $\mathcal{B}_1$ and $\mathcal{B}_2$ are satisfied for the stochastic integrals $\Delta_n(\vartheta_0)$ and $\Lambda_n(u)$ under $\mathcal{H}_0$ and $\mathcal{H}_u$ with some sequence $\varepsilon_n \to 0$, where $\varepsilon_n = O(n^{-q/2})$ for some $q > 0$.

$\mathcal{D}_3$. There exists some functions $f_j(x),\ x \in \mathbb{A}_n\, j = 0, 1, 2$ not depending on ϑ such that $S(\vartheta, x) \geq f_0(x), |S^{(j)}(\vartheta, x)| \leq f_j(x), j = 0, 1, 2$ for all $x \in \mathbb{A}_n$ and all ϑ in a right neighborhood of ϑ_0. We suppose also that (for $k = 2, 3, 4$)

$$\varphi_n^k \int_{\mathbb{A}_n} \frac{|f_1(x)|^k}{f_0(x)^{k-1}}\, \mathrm{d}x = O(\varepsilon_n^{k-2}),\ \varphi_n^4 \int_{\mathbb{A}_n} \frac{f_2(x)^2}{f_0(x)}\, \mathrm{d}x = O(\varepsilon_n^2).$$

$\mathcal{D}_4$. For any sequence $\{k_n\}$ of integers satisfying $\frac{\varphi_n}{\varphi_{k_n}} = 1 + O(\varepsilon_n^2)$ and $\mathbb{A}_n \subseteq \mathbb{A}_{k_n}$, we have (for $j = 1, 2$)

$$\varphi_n^{2j} \int_{\mathbb{A}_{k_n} - \mathbb{A}_n} \frac{f_j(x)^2}{f_0(x)}\, \mathrm{d}x = O(\varepsilon_n^{2j}),\ \varphi_n^4 \int_{\mathbb{A}_{k_n} - \mathbb{A}_n} \frac{f_1(x)^4}{f_0(x)^3}\, \mathrm{d}x = O(\varepsilon_n^4).$$

$\mathcal{D}_5$. For any sequence $\{k_n\}$ of integers satisfying $\frac{\varphi_n}{\varphi_{k_n}} = 1 + O(\varepsilon_n)$ and $\mathbb{A}_n \subseteq \mathbb{A}_{k_n}$, we have (for $j = 1, 2$)

$$\varphi_n^{2j} \int_{\mathbb{A}_{k_n} - \mathbb{A}_n} \frac{f_j(x)^2}{f_0(x)}\, \mathrm{d}x = O(\varepsilon_n^{2j-1}),\ \varphi_n^3 \int_{\mathbb{A}_{k_n} - \mathbb{A}_n} \frac{|f_1(x)|^3}{f_0(x)^2}\, \mathrm{d}x = O(\varepsilon_n^2).$$

Remark 3.1. In the condition $\mathcal{D}_1$, we can suppose without loss of generality that the normalizing factor $\varphi_n = C\, n^{-p}$, for some $C > 0$. Because one can write $\vartheta_u = \vartheta_0 + \varphi_n\, u = \vartheta_0 + C\, n^{-p}\, u_n$, where $u_n \to u$ as $n \to \infty$.

Example 1. Let $X^{(n)}$ be a realization of a Poisson process on the set $\mathbb{A}_n = [\,0, n]$ with positive intensity function $S(\vartheta, x) = \vartheta S(x) + \lambda$ (amplitude parameter) or $S(\vartheta, x) = S(\vartheta + x) + \lambda$ (phase parameter), where $S(\cdot)$ is a two times differentiable periodic function and $\lambda > 0$ (dark current) is a known parameter. In both cases the conditions $\mathcal{D}_3 - \mathcal{D}_5$ are satisfied with $\varphi_n \sim C\, n^{-1/2}$ for some $C > 0$ and $\varepsilon_n = n^{-1/2}$. For the frequency modulation model $S(\vartheta, x) = S(\vartheta\, x) + \lambda$ we have $\varphi_n \sim C\, n^{-3/2}$ and $\varepsilon_n = n^{-1/2}$.

The condition $\mathcal{D}_2$ allows us to write the powers as follows:

$$\begin{aligned}
\beta_n\,(\vartheta_u, \phi_n^*) &= \mathcal{N}\,(\Delta) + \mathrm{n}(\Delta)\, E_1^{(n)}(u, z_\alpha)\, \varepsilon_n + \mathrm{n}(\Delta)\, E_2^{(n)}(u, z_\alpha)\, \varepsilon_n^2 + O(\varepsilon_n^3) \\
\beta_n\left(\vartheta_u, \tilde{\phi}_n\right) &= \mathcal{N}\,(\Delta) + \mathrm{n}(\Delta)\, E_1^{(n)}(u, z_\alpha)\, \varepsilon_n + \mathrm{n}(\Delta)\, F_2^{(n)}(u, z_\alpha)\, \varepsilon_n^2 + O(\varepsilon_n^3), \quad (5)
\end{aligned}$$

where $\Delta = \Delta(u) = u - z_\alpha$. The equality of the second terms in the right hand sides follows from the second order efficiency of the score test ϕ_n^* which is given by

$$E_1^{(n)}(u, z_\alpha)\, \varepsilon_n = \frac{u(z_\alpha - 2u)}{6}\, \gamma_{3,n} + \frac{\varphi_n^3\, u^2}{2} \int_{\mathbb{A}_n} \frac{S^{(1)}(\vartheta_0, x)\, S^{(2)}(\vartheta_0, x)}{S(\vartheta_0, x)}\, \mathrm{d}x.$$

See [5] and [7]. The deficiency arises from the following nonequal terms $E_2^{(n)}(u, z_\alpha)$ and $F_2^{(n)}(u, z_\alpha)$ (of orders $O(1)$):

$$\begin{aligned}
E_2^{(n)}(u, z_\alpha)\, \varepsilon_n^2 &= \frac{\gamma_{3,n}(u)}{6}(1 - a_n^2)(a_n - \Delta)\Delta - \frac{(a_n - \Delta)^2}{2}\Delta + \\
&\quad + \frac{\gamma_{4,n}(u)}{4!} H_3(\Delta) + \frac{\gamma_{3,n}^2(u)}{72} H_5(\Delta) \\
F_2^{(n)}(u, z_\alpha)\, \varepsilon_n^2 &= A_n - a_n - \frac{\gamma'_{3,n}(u) - \gamma_{3,n}(u)}{6}(1 - \Delta^2) - \frac{(A_n - \Delta)^2}{2}\Delta + \\
&\quad + \frac{\gamma'_{4,n}(u)}{4!} H_3(\Delta) + \frac{\gamma'^2_{3,n}(u)}{72} H_5(\Delta).
\end{aligned}$$

Below we give the terms appeared in the above right hand sides and in what follows.

$$m_n(u) = \mathbf{E}_{\vartheta_u}\Delta_n(\vartheta_0) = \varphi_n \int_{\mathbb{A}_n} \frac{S^{(1)}(\vartheta_0, x)}{S(\vartheta_0, x)} \left(S(\vartheta_u, x) - S(\vartheta_0, x)\right)\, \mathrm{d}x,$$

$$\eta_n^2(u) = \mathbf{E}_{\vartheta_u}\left(\Delta_n(\vartheta_0) - m_n(u)\right)^2 = \varphi_n^2 \int_{\mathbb{A}_n} \frac{S^{(1)}(\vartheta_0, x)^2}{S(\vartheta_0, x)^2} S(\vartheta_u, x)\, \mathrm{d}x,$$

$$\gamma_{r,n}(u) = \frac{\varphi_n^r}{\eta_n^r(u)} \int_{\mathbb{A}_n} \frac{S^{(1)}(\vartheta_0, x)^r}{S(\vartheta_0, x)^r} S(\vartheta_u, x)\, \mathrm{d}x \qquad r = 3, 4, \tag{6}$$

and the corresponding terms for $\Lambda_n(u)$ under the local alternative

$$\mu_n(u) = \mathbf{E}_{\vartheta_u}\Lambda_n(u) = \int_{\mathbb{A}_n} \left(\ln \frac{S(\vartheta_u, x)}{S(\vartheta_0, x)} S(\vartheta_u, x) - S(\vartheta_u, x) + S(\vartheta_0, x)\right) \mathrm{d}x,$$

$$\sigma_n^2(u) = \mathbf{E}_{\vartheta_u}\left(\Lambda_n(u) - \mu_n(u)\right)^2 = \int_{\mathbb{A}_n} \left(\ln \frac{S(\vartheta_u, x)}{S(\vartheta_0, x)}\right)^2 S(\vartheta_u, x)\, \mathrm{d}x \tag{7}$$

$$\gamma'_{r,n}(u) = \frac{1}{\sigma_n(u)^r} \int_{\mathbb{A}_n} \left(\ln \frac{S(\vartheta_u, x)}{S(\vartheta_0, x)}\right)^r S(\vartheta_u, x)\, \mathrm{d}x \qquad r = 3, 4.$$

We introduce also the normalized thresholds

$$a_n = \frac{m_n(u) - c_n}{\eta_n(u)}, \qquad A_n = \frac{\mu_n(u) - b_n(u)}{\sigma_n(u)}.$$

Since the investigation is at accuracy level $O(\varepsilon_n^3)$, by Theorem 1.1, part (ii) we can write:

$$c_n = z_\alpha + \frac{\gamma_{3,n}}{3!} H_2(z_\alpha) + \frac{\gamma_{4,n}}{4!} H_3(z_\alpha) + \frac{\gamma_{3,n}^2}{72} H_5(z_\alpha), \tag{8}$$

where

$$\gamma_{r,n} = \varphi_n^r \int_{\mathbb{A}_n} \frac{S^{(1)}(\vartheta_0, x)^r}{S(\vartheta_0, x)^{r-1}}\, \mathrm{d}x, \qquad r = 3, 4.$$

With this choice of c_n, the probability of the error of the first kind of ϕ_n^* is equal to $\alpha + O(\varepsilon_n^3)$. Similarly we have

$$b_n(u) = \mu_n + \sigma_n \left(z_\alpha + \frac{\gamma'_{3,n}}{6} H_2(z_\alpha) + \frac{\gamma'_{4,n}}{4!} H_3(z_\alpha) + \frac{\gamma'^2_{3,n}}{72} H_5(z_\alpha)\right),$$

up to order $O(\varepsilon_n^3)$, where

$$\mu_n = \mathbf{E}_{\vartheta_0}\Lambda_n(u) = \int_{\mathbb{A}_n}\left(\ln\frac{S(\vartheta_u,x)}{S(\vartheta_0,x)} - \frac{S(\vartheta_u,x)}{S(\vartheta_0,x)} + 1\right) S(\vartheta_0,x)\,\mathrm{d}x,$$

$$\sigma_n^2 = \mathbf{E}_{\vartheta_0}\left(\Lambda_n(u)-\mu_n\right)^2 = \int_{\mathbb{A}_n}\left(\ln\frac{S(\vartheta_u,x)}{S(\vartheta_0,x)}\right)^2 S(\vartheta_0,x)\,\mathrm{d}x,$$

$$\gamma'_{r,n} = \frac{1}{\sigma_n^r}\int_{\mathbb{A}_n}\left(\ln\frac{S(\vartheta_u,x)}{S(\vartheta_0,x)}\right)^r S(\vartheta_0,x)\,\mathrm{d}x, \quad r=3,4.$$

The main result is the following theorem.

Theorem 3.1. *Let the conditions $\mathcal{D}_1-\mathcal{D}_6$ be fulfilled. Then for any $u>0$,*

$$d = \frac{q}{p\,u}\lim_{n\to\infty}\left(\varepsilon_n^2\, n^q\left(F_2^{(n)}-E_2^{(n)}\right)\right),$$

provided that the limit exists.

If $\varepsilon_n = n^{-q/2}$, then

$$d = \frac{q\,r(u)}{u\,p\,\mathrm{n}(u-z_\alpha)}$$

where $r(u)$ is the power loss of ϕ_n^* with respect to $\tilde{\phi}_n$, given by

$$r(u) = \frac{u^3\,\mathrm{n}(u-z_\alpha)}{8}\lim_{n\to\infty}\left(\varepsilon_n^{-2}\,J_n\right),$$

for any $u>0$ (see [7]) and

$$J_n = \varphi_n^4\int_{\mathbb{A}_n}\frac{\left(S^{(1)}(\vartheta_0,x)^2 - S(\vartheta_0,x)\,S^{(2)}(\vartheta_0,x)\right)^2}{S(\vartheta_0,x)^3}\mathrm{d}x -$$
$$-\left(\varphi_n^3\int_{\mathbb{A}_n}\frac{S^{(1)}(\vartheta_0,x)\left(S^{(1)}(\vartheta_0,x)^2 - S(\vartheta_0,x)\,S^{(2)}(\vartheta_0,x)\right)}{S(\vartheta_0,x)^2}\mathrm{d}x\right)^2.$$

Proof. Here we shall sketch the proof of the theorem and omit the details. As mentioned earlier, the problem of asymptotic deficiency is at the accuracy level $O(\varepsilon_n^3)$, hence instead of $\beta_{k_n}\left(\vartheta_u,\phi_{k_n}^*\right) = \beta_n(\vartheta_u,\tilde{\phi}_n)$, it is sufficient to consider

$$\beta_{k_n}\left(\vartheta_u,\phi_{k_n}^*\right) = \beta_n(\vartheta_u,\tilde{\phi}_n) + O(\varepsilon_n^3), \tag{9}$$

as $n\to\infty$. Now we show that this implies

$$\frac{\varphi_n}{\varphi_{k_n}} = 1 + O(\varepsilon_n^2) \tag{10}$$

as $n\to\infty$. If we consider just the first term in (5), we have $\beta_n(\vartheta_u,\tilde{\phi}_n) = \mathcal{N}(u-z_\alpha)+O(\varepsilon_n)$, and the power of $\phi_{k_n}^*$ at the local alternative $\vartheta_u = \vartheta_0+\varphi_n\,u =$

$\vartheta_0+\varphi_{k_n}\frac{\varphi_n}{\varphi_{k_n}}\,u$, can be written as $\beta_{k_n}\left(\vartheta_u,\phi^*_{k_n}\right)=\mathcal{N}\left(\frac{\varphi_n}{\varphi_{k_n}}\,u-z_\alpha\right)+O(\varepsilon_{k_n})$. Hence (9) implies that $\mathcal{N}\left(\frac{\varphi_n}{\varphi_{k_n}}\,u-z_\alpha\right)=\mathcal{N}\left(u-z_\alpha\right)+O(\varepsilon_n)$, from which we get

$$\frac{\varphi_n}{\varphi_{k_n}}=1+O(\varepsilon_n). \tag{11}$$

Again using (5) we can write

$$\beta_{k_n}\left(\vartheta_u,\phi^*_{k_n}\right)=\mathcal{N}\left(u_n-z_\alpha\right)+Q_{k_n}(u_n)\,n(u_n-z_\alpha)+O(\varepsilon_n^2),$$

where $Q_n(u)=E_1^{(n)}(u,z_\alpha)\,\varepsilon_n$ and $u_n=\frac{\varphi_n}{\varphi_{k_n}}u$. Hence

$$Q_{k_n}(u_n)=\frac{u_n\,\left(z_\alpha-2\,u_n\right)}{6}\,\gamma_{3,k_n}+\frac{\varphi_{k_n}^3\,u_n^2}{2}\int_{\mathbb{A}_{k_n}}\frac{S^{(1)}(\vartheta_0,x)\,S^{(2)}(\vartheta_0,x)}{S(\vartheta_0,x)}\,\mathrm{d}x.$$

From $\mathcal{D}_3$, $\mathcal{D}_5$ and (11) we obtain

$$\gamma_{3,k_n}=\varphi_{k_n}^3\int_{\mathbb{A}_{k_n}}\frac{S^{(1)}\left(\vartheta_0,x\right)^3}{S\left(\vartheta_0,x\right)^2}\,\mathrm{d}x=\frac{\varphi_{k_n}^3}{\varphi_n^3}\,\varphi_n^3\int_{\mathbb{A}_n}\frac{S^{(1)}\left(\vartheta_0,x\right)^3}{S\left(\vartheta_0,x\right)^2}\,\mathrm{d}x+$$
$$\frac{\varphi_{k_n}^3}{\varphi_n^3}\,\varphi_n^3\int_{\mathbb{A}_{k_n}-\mathbb{A}_n}\frac{S^{(1)}\left(\vartheta_0,x\right)^3}{S\left(\vartheta_0,x\right)^2}\,\mathrm{d}x=\left(1+O(\varepsilon_n)\right)\gamma_{3,n}+O(\varepsilon_n^2)=\gamma_{3,n}+O(\varepsilon_n^2).$$

Note that $\gamma_{3,n}=O(\varepsilon_n)$ by $\mathcal{D}_3$. In a similar way from (11), $\mathcal{D}_5$ and the Cauchy-Schwartz inequality we have

$$\varphi_{k_n}^3\int_{\mathbb{A}_{k_n}}\frac{S^{(1)}(\vartheta_0,x)\,S^{(2)}(\vartheta_0,x)}{S(\vartheta_0,x)}\,\mathrm{d}x=\varphi_n^3\int_{\mathbb{A}_n}\frac{S^{(1)}(\vartheta_0,x)\,S^{(2)}(\vartheta_0,x)}{S(\vartheta_0,x)}\,\mathrm{d}x+O(\varepsilon_n^2).$$

Therefore $Q_{k_n}(u_n)=Q_n(u)+O(\varepsilon_n^2)$ and consequently from (11) we have

$$\beta_{k_n}\left(\vartheta_u,\phi^*_{k_n}\right)=\mathcal{N}\left(\Delta\right)+u\left(\frac{\varphi_n}{\varphi_{k_n}}-1\right)\,n(\Delta)+Q_n(u)\,n(\Delta)+O(\varepsilon_n^2),$$

where $\Delta=u-z_\alpha$. Now using (5) and (9), we can write

$$\beta_n\left(\vartheta_u,\tilde{\phi}_n\right)=\mathcal{N}\left(\Delta\right)+u\left(\frac{\varphi_n}{\varphi_{k_n}}-1\right)\,n(\Delta)+Q_n(u)\,n(\Delta)+O(\varepsilon_n^2)=$$
$$\mathcal{N}\left(\Delta\right)+n(\Delta)\,E_1^{(n)}(u,z_\alpha)\,\varepsilon_n+n(\Delta)\,F_2^{(n)}(u,z_\alpha)\,\varepsilon_n^2+O(\varepsilon_n^3)=$$
$$\mathcal{N}\left(\Delta\right)+Q_n(u)\,n(\Delta)+O(\varepsilon_n^2),$$

which proves (10). Using (10) and $\mathcal{D}_4$ we have

$$E_1^{(k_n)}(u_n,z_\alpha)\,n(\frac{\varphi_n}{\varphi_{k_n}}u-z_\alpha)\,\varepsilon_{k_n}=E_1^{(n)}(u,z_\alpha)\,n(u-z_\alpha)\,\varepsilon_n+O(\varepsilon_n^3)$$
$$E_2^{(k_n)}(u_n,z_\alpha)\,n(\frac{\varphi_n}{\varphi_{k_n}}u-z_\alpha)\,\varepsilon_{k_n}^2=E_2^{(n)}(u,z_\alpha)\,n(u-z_\alpha)\,\varepsilon_n^2+O(\varepsilon_n^3).$$

Here we used the fact that $\gamma_{3,k_n}(u) = \gamma_{3,n}(u) + O(\varepsilon_n^3)$ and $\gamma_{4,k_n}(u) = \gamma_{4,n}(u) + O(\varepsilon_n^3)$, which follows from $\mathcal{D}_4$. Therefore

$$\beta_{k_n}\left(\vartheta_u, \phi^*_{k_n}\right) = \mathcal{N}(\Delta) + \left(\frac{\varphi_n}{\varphi_{k_n}} - 1\right) u\, n(\Delta) + E_1^{(n)}(u, z_\alpha)\, n(\Delta)\, \varepsilon_n + \\ + E_2^{(n)}(u, z_\alpha)\, n(\Delta)\, \varepsilon_n^2 + O(\varepsilon_n^3)$$

and consequently (9) leads to

$$\left(\frac{\varphi_n}{\varphi_{k_n}} - 1\right) u = \left(F_2^{(n)}(u, z_\alpha) - E_2^{(n)}(u, z_\alpha)\right) \varepsilon_n^2 + O(\varepsilon_n^3).$$

Since $\varphi_n = Cn^{-p}$ and $\varepsilon_n = O(n^{-q/2})$ for some $C > 0$ and $p > 0$ we obtain

$$\left(\left(\frac{k_n}{n}\right)^p - 1\right) u = \left((F_2^{(n)}(u, z_\alpha) - E_2^{(n)}(u, z_\alpha))\right) \varepsilon_n^2 + O(\varepsilon_n^3).$$

Now using the mean value theorem

$$k_n^q - n^q = \frac{q}{p\, u}\, \xi_n \left((F_2^{(n)}(u, z_\alpha) - E_2^{(n)}(u, z_\alpha))\right) n^q\, \varepsilon_n^2 + O(\varepsilon_n^{-q/2}),$$

where the intermediate point $\xi_n \to 1$ as $n \to \infty$. This proves the theorem.

Example 2. We observe the Poisson process $X^{(n)}$ with intensity function $S(\vartheta, x) = e^{\sin(\vartheta\, r)}$, $\vartheta > 0$, $x \in \mathbb{A}_n = \{x : |x| \le n\} \subset \mathbb{R}^2$, $\quad n = 1, 2, ...$, the point $x = (x_1, x_2)$ and $r = |x| = \sqrt{x_1^2 + x_2^2}$. We test $\mathcal{H}_0 : \vartheta = \vartheta_0$ against the one sided alternative $\mathcal{H}_1 : \vartheta > \vartheta_0$. It can be shown that the conditions of the theorem are fulfilled, $\varphi_n \sim C\, n^{-2}$ for some $C > 0$ and $\varepsilon_n = n^{-1}$, *i.e.*, $p = q = 2$. Hence we will obtain

$$\lim_{n\to\infty} d_n = \lim_{n\to\infty} \left(k_n^2 - n^2\right) = \frac{r(u)}{u\, n(u - z_\alpha)},$$

where the power loss is given by $r(u) = \frac{1}{8\,\pi}\, u^3\, n(u - z_\alpha)\; C_0^{-3}\, (4\, C_0\, C_1 - C_2^2)$, for any $u > 0$ and some constants C_0, C_1 and C_2 (see [7]).

Example 3. Suppose that we observe a realization $X^{(n)}$ of a Poisson process on the set $\mathbb{A}_n = [\,0, n]$, $n = 1, 2, \cdots$ with the intensity function $S(\vartheta, x) = \vartheta\, S(x) + \lambda, \quad \vartheta > 0$ where λ is a known positive constant (dark-current) and $S(x)$ is a known, nonconstant, differentiable and periodic function with period $\tau > 0$. The intensity function $S(\vartheta, x)$ is supposed to be positive in a right neighborhood of ϑ_0 and all x. The conditions of the theorem are fulfilled and $\varphi_n \sim C\, n^{-1/2}$ for some $C > 0$ and $\varepsilon_n = n^{-1/2}$. In this case $p = 1/2$ and $q = 1$. The asymptotic deficiency is equal to

$$\lim_{n\to\infty} d_n = \lim_{n\to\infty} (k_n - n) = \frac{2\, r(u)}{u\, n(u - z_\alpha)},$$

where $r(u) = \frac{1}{8}\, u^3\, n(u - z_\alpha)\; D_0^{-3}\, (D_0\, D_1 - D_2^2)$, for any $u > 0$ and some constants D_0, D_1 and D_2 (see [7]).

References

1. Bening, V.E. (2000). *Asymptotic Theory of Testing Statistical Hypotheses*. VSP, Utrecht.
2. Dachian, S., Kutoyants, Yu.A. (2006) *Hypotheses testing: Poisson versus self-exciting*, Scandinavian J. of Statistics, **33**: 391-408.
3. Daley, D.J., Vere-Jones, D. (1988). *An introduction to the theory of point processes*. Springer, New York.
4. Davies, R. (1977). *Testing the hypothesis that a point process is Poisson*, Adv. in Appl. Probab., **9**: 724-746.
5. Fazli, Kh. (2007). *Second order efficient test for inhomogeneous Poisson Processes*, Statistical Inference for Stochastic Processes. **10**: 181-208.
6. Fazli, Kh. (2009). *Hypotheses Testing for a Multidimensional Parameter of Inhomogeneous Poisson Processes*, Statistics, **42**: 407-428.
7. Fazli, Kh. *Power Loss for Inhomogeneous Poisson Processes*, to appear.
8. Ingster Yu. I., Kutoyants Yu. A. (2007) *Nonparametric hypothesis testing for intensity of the Poisson process*, Mathematical Methods of Statistics, **16**: 217-245.
9. Kutoyants, Yu.A. (1998). *Statistical Inference for Spatial Poisson Processes*. Springer, New York.
10. Pitman, E.J.G. (1948). *Lecture Notes on Nonparametric Statistical Inferences*. Lectures given for the University of North Carolina, Institue of Statistics.
11. Roussas, G.G. (1972). *Contiguity of Probability Measure*. Cambridge University Press, Cambridge.

A Note On Computer Simulation Of Geometric Stable Random Vectors

M. Firouzi, H. Zare and A. Mohammadpour

Department of Statistics, Amirkabir University of Technology

In this paper, a modified representation of Kozubowski and Panorska's method, 1999, for simulating geometric stable random vectors on a computer is presented. The method is based on a representation for these vectors as mixtures of α-stable random vectors. A S-PLUS function for simulating such vectors and a simulation study for compare with the previous method are given.

Keywords: Geometric stable random vectors, Spectral measure, Simulation.

1. Introduction

Geometric stable distributions provide good approximations for normalized sums of i.i.d. random variables, where the number of terms in the summation has a geometric distribution with the parameter p that p converging to zero and is independent from the terms. Geometric stable distributions have many applications in financial modeling (see, e.g., [1])

In [2] has introduced a representation for α-stable random vectors as a linear combination of vector multiples of independent totally skewed to the right stable random variables with $0 < \alpha < 2$. On the basis of this representation, in [3] has introduced a method for simulating geometric stable random vectors.

In this paper, we modified the Kozubowski and Panorska's simulating method and represented a faster algorithm for simulation. In section 2, we recall the definition of geometric stable random vectors and present a representation of them in terms of α-stable random vectors. Modified method and its corresponding simulation algorithm described in section 3. Eventually in a simulation study we compare modified algorithm with the previous one.

2. Geometric stable random vectors

Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ be a sequence of i.i.d. random vectors in $\mathbb{R}^d$. If s_p be a geometric random variable with parameter $0 < p < 1$ and the following probability function

$$P(s_p = n) = p(1-p)^{n-1}, \quad n = 1, 2, \ldots,$$

independent of $\{\boldsymbol{Y}_i\}$, then we have the following definition for a geometric stable random vector.

Definition 2.1. A random vector $\boldsymbol{X}$ is said to have a geometric stable distribution if it has a domain of (geometric) attraction, i.e., there is a geometric random variable s_p independent of sequence of i.i.d. random vectors $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$, and $a(p) > 0, \boldsymbol{b}(p) \in \mathbb{R}^d$ such that

$$a(p) \sum_{i=1}^{s_p} (\boldsymbol{Y}_i + \boldsymbol{b}(p)) \stackrel{d}{\Rightarrow} \boldsymbol{X}, \quad as\ p \to 0.$$

The best way to describe the distribution of a geometric stable random vector $\boldsymbol{X}$ is by its characteristic function. A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is said to be a geometric stable random vector in $\mathbb{R}^d$ if and only if there exists a finite measure Γ on the unit sphere $\mathbb{S}_d$ of $\mathbb{R}^d$, and a vector $\boldsymbol{\mu} \in \mathbb{R}^d$, such that

$$\Phi_{\boldsymbol{X}}(\boldsymbol{t}) = \Big(1 + \int_{S^d} \psi(< \boldsymbol{t}, \boldsymbol{s} >)\Gamma(d\boldsymbol{s}) - i < \boldsymbol{t}, \boldsymbol{\mu} > \Big)^{-1}, \tag{1}$$

where

$$\psi(u) = \begin{cases} |u|^\alpha\{1 - i \tan\frac{\pi\alpha}{2}\text{sign}(u)\}, \ \alpha \neq 1, \\ |u|\{1 + i\frac{2}{\pi}\text{sign}(u)\log|u|\}, \ \alpha = 1. \end{cases}$$

The vector $\boldsymbol{X}$ is said to have spectral representation $(\Gamma, \boldsymbol{\mu})$, and denote as $\boldsymbol{X} \sim GS(\Gamma, \boldsymbol{\mu})$. The measure Γ can approximated by a discrete measure with a finite number of point masses, i.e.,

$$\Gamma(\cdot) = \sum_{j=1}^{n} \gamma_j \delta_{\boldsymbol{s}_j}(\cdot). \tag{2}$$

where weight $\gamma_j > 0$, and $\delta_{\boldsymbol{s}_j}$'s are the point masses at $\boldsymbol{s}_j \in \mathbb{S}_d, j = 1, \ldots, n$. In this case, 1 reduce to

$$\Phi_{\boldsymbol{X}}(\boldsymbol{t}) = \Big(1 + \sum_{j=1}^{n} \psi(< \boldsymbol{t}, \boldsymbol{s}_j >)\gamma_j - i < \boldsymbol{t}, \boldsymbol{\mu} > \Big)^{-1}. \tag{3}$$

The following theorem, which appeared in [1], represents geometric stable random vectors as mixture representations of α-stable random vectors.

Theorem 2.1. *Let $\boldsymbol{X}$ be a geometric stable random vector with index α, spectral measure Γ, and a location parameter $\boldsymbol{\mu}$, and let $\boldsymbol{Y}$ be an α-stable random vector with the same spectral measure and location parameter $\mathbf{0}$. Let Z be an exponentially distributed random variable with mean 1, independent of $\boldsymbol{Y}$. Then*

$$\boldsymbol{X} \stackrel{d}{=} \begin{cases} Z^{\frac{1}{\alpha}}\boldsymbol{Y} + \boldsymbol{\mu}Z, & \alpha \neq 1, \\ Z\boldsymbol{Y} + (\frac{2}{\pi}Z\log Z)\boldsymbol{g} + \boldsymbol{\mu}Z, & \alpha = 1, \end{cases} \tag{4}$$

where

$$\boldsymbol{g} = (g_1, g_2, \ldots, g_d), \quad g_k = \int_{S^d} \boldsymbol{s}_k \Gamma(d\boldsymbol{s}).$$

3. simulation

Let V be a stable random variable with the following characteristic function

$$\phi_V(t) = \begin{cases} \exp\{-\gamma^\alpha |t|^\alpha \{1 - i\beta \text{sign}(t)\tan\frac{\pi\alpha}{2}\} + i\delta t\}, & \alpha \neq 1, \\ \exp\{-\gamma|t|\{1 + i\beta\frac{2}{\pi}\text{sign}(u)\log|t|\} + i\delta t\}, & \alpha = 1, \end{cases}$$

where $\alpha \in (0, 2]$ is the index of stability, $\beta \in [-1, 1]$ is the skewness parameter, $\gamma > 0$ is the scale parameter and $\delta \in \mathbb{R}$ is the location parameter. We will use the symbol $S(\alpha, \beta, \gamma, \delta)$ for the α-stable distributions. If $\gamma = 1$ and $\delta = 0$ the distribution is standard and we denote by $S(\alpha, \beta)$. If $V \sim S(\alpha, 1)$, we will say V is totally skewed to the right.

The following representation for stable random vector $\boldsymbol{Y}$ with a location parameter $\mathbf{0}$ and discrete spectral measure 2, appeared in [2].

$$\boldsymbol{Y} \stackrel{d}{=} \begin{cases} \sum_{j=1}^n \gamma_j^{\frac{1}{\alpha}} V_j \boldsymbol{s}_j, & \alpha \neq 1, \\ \sum_{j=1}^n \gamma_j (V_j + \frac{2}{\pi}\log\gamma_j)\boldsymbol{s}_j, & \alpha = 1, \end{cases} \tag{5}$$

where $V_j \sim S(\alpha, 1), j = 1, \ldots, n$.

The following result, that obtained from 4 and 5, was given in [3], and gives a representation of a geometric stable random vector in terms of vector multiples of independent totally skewed to the right stable random variables.

Theorem 3.1. *Let $\boldsymbol{X} \sim GS(\Gamma, \boldsymbol{\mu})$, where Γ is as in 2, and $0 < \alpha < 2$. Let Z be a exponential random variable with mean 1, and let $V_1, \ldots, V_n$ be i.i.d. $S(\alpha, 1)$ random variables, independent of Z. Then*

$$\boldsymbol{X} \stackrel{d}{=} Z\left[Z^{\frac{1}{\alpha}-1}\sum_{j=1}^n \gamma_j^{\frac{1}{\alpha}}(V_j + I_{\{1\}}(\alpha)\frac{2}{\pi}\log(\gamma_j Z))\boldsymbol{s}_j + \boldsymbol{\mu}\right], \tag{6}$$

where I denote indicator function.

We call this method M1 and obtain the following algorithm to simulating geometric stable random vectors.

Algorithm 1.

- Generate n independent random variables $V_j, j = 1, \ldots, n$ from $S(\alpha, 1)$.
- Generate a standard exponential random variable Z, independent of V_j's.
- Compute

$$\boldsymbol{X} = Z^{\frac{1}{\alpha}}(A\boldsymbol{V} + \boldsymbol{b}) + \boldsymbol{\Delta}$$

where $\boldsymbol{V} = (V_1 + I_{\{1\}}(\alpha)\frac{2}{\pi}\log Z, \ldots, V_n + I_{\{1\}}(\alpha)\frac{2}{\pi}\log Z)', A = (\gamma_1^{\frac{1}{\alpha}}\boldsymbol{s}_1, \ldots, \gamma_n^{\frac{1}{\alpha}}\boldsymbol{s}_n), \boldsymbol{b} = I_{\{1\}}(\alpha)\frac{2}{\pi}A\boldsymbol{c}, \boldsymbol{c} = (\log\gamma_1, \ldots, \log\gamma_n)'$, and $\boldsymbol{\Delta} = (z\mu_1, \ldots, z\mu_d)'$. The following S-PLUS function can be use for generate N pseudo stable random vectors based on M1 method.

Function name: rmvgstab1

```
function(N, alpha, LG, Ga, mu) {
# LG = location of point masses on the unit sphere
# Ga = vector of point masses, (n positive elements)
  n <- dim(LG)[1]  # n = number of point masses
  d <- dim(LG)[2]  # d = dimension of random vectors (rv)
  Z1 <- matrix(rexp(N), nrow = N, ncol = n)
  Z <- Z1[, c(1:d)]
  V <- matrix(rstab(N * n, alpha, 1) + if(alpha == 1) (2/pi) * log(Z1)
         else tan((pi * alpha)/2), ncol = n)
  Ga1 <- matrix(Ga, nrow = n, ncol = d)
  A <- LG * Ga1^(1/alpha)
  b <- matrix(if(alpha == 1) (2/pi) * log(Ga) %*% A
          else 0, nrow = N, ncol = d, byrow = T) * Z^(1/alpha)
  mu <- matrix(mu, nrow = N, ncol = d, byrow = T) * Z + b
  (V %*% A) * Z^(1/alpha) + mu
}
```

When Γ is a symmetric discrete measure on $\mathbb{S}_d$, there are an even number of point masses and it be assumed that $\gamma_j = \gamma_{j+m}, \boldsymbol{s}_j = -\boldsymbol{s}_{j+m}$, for $j = 1, \ldots, m$, and $n = 2m$. Then 3, reduces to

$$\Phi_{\boldsymbol{X}}(\boldsymbol{t}) = \Big(1 + 2\sum_{j=1}^{m}\psi(<\boldsymbol{t}, \boldsymbol{s}_j>)\gamma_j - i<\boldsymbol{t}, \boldsymbol{\mu}>\Big)^{-1}.$$

In this case 5, takes the following form

$$\boldsymbol{Y} \stackrel{d}{=} \sum_{j=1}^{m}(2\gamma_j)^{\frac{1}{\alpha}}W_j\boldsymbol{s}_j,$$

where $W_j \sim S(\alpha, 0), j = 1, \ldots, m$, (see, [2]).

If the spectral measure has some symmetric point masses, then simulation time can be reduced. The following lemma that appeared in [4] and next theorem give a representation of a nonsymmetric geometric stable random vector with a discrete spectral measure that can decrease the simulation times.

Lemma 3.1. *Let* $0 < \alpha < 2, \gamma_1, \ldots, \gamma_n > 0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathbb{S}_d, \gamma_j = \gamma_{j+m}, \boldsymbol{s}_j = -\boldsymbol{s}_{j+m}$, *for* $j = n_1 + 1, \ldots, n_1 + m, 0 \le n_1 \le n, n - n_1 = 2m$, *and* $V_1, \ldots, V_{n_1}$ *be independently distributed with* $S(\alpha, 1)$, $W_{n_1+1}, \ldots, W_{n_1+m}$ *be independently distributed with* $S(\alpha, 0)$. *If*

$\boldsymbol{Y}$ *be an* α*-stable random vector with discrete spectral measure 2, and location parameter* $\mathbf{0}$*. Then*

$$\boldsymbol{Y} \stackrel{d}{=} \sum_{j=1}^{n_1} \gamma_j^{\frac{1}{\alpha}} (V_j + I_{\{1\}}(\alpha)\frac{2}{\pi}\log\gamma_j)\boldsymbol{s}_j + \sum_{j=n_1+1}^{n_1+m} (2\gamma_j)^{\frac{1}{\alpha}} W_j \boldsymbol{s}_j. \tag{7}$$

Theorem 3.2. *Let* $\boldsymbol{X} \sim GS(\Gamma, \boldsymbol{\mu})$*, where* Γ *is as in 2, and* $0 < \alpha < 2$*. Let* Z *be a exponential random variable with mean 1,* $V_1, \ldots, V_{n_1}$ *be i.i.d.* $S(\alpha, 1)$ *random variables and* $W_{n_1+1}, \ldots, W_{n_1+m}$ *be i.i.d.* $S(\alpha, 0)$ *random variables, independent of* Z *and* V_j*'s. Then*

$$\begin{aligned}\boldsymbol{X} \stackrel{d}{=} Z\Big[Z^{\frac{1}{\alpha}-1}\Big[\sum_{j=1}^{n_1} \gamma_j^{\frac{1}{\alpha}}(V_j + I_{\{1\}}(\alpha)\frac{2}{\pi}\log(\gamma_j Z))\boldsymbol{s}_j \\ + \sum_{j=n_1+1}^{n_1+m} (2\gamma_j)^{\frac{1}{\alpha}}(W_j + I_{\{1\}}(\alpha)\frac{\log Z}{\pi})\boldsymbol{s}_j\Big] + \boldsymbol{\mu}\Big]\end{aligned}$$

Proof. 1. Case of $\alpha \neq 1$: From 4, and 7, we have

$$\begin{aligned}\boldsymbol{X} &\stackrel{d}{=} Z^{\frac{1}{\alpha}}\boldsymbol{Y} + \boldsymbol{\mu}Z \\ &\stackrel{d}{=} Z^{\frac{1}{\alpha}}\Big[\sum_{j=1}^{n_1} \gamma_j^{\frac{1}{\alpha}} V_j \boldsymbol{s}_j + \sum_{j=n_1+1}^{n_1+m} (2\gamma_j)^{\frac{1}{\alpha}} W_j \boldsymbol{s}_j\Big] + \boldsymbol{\mu}Z \\ &\stackrel{d}{=} Z\Big[Z^{\frac{1}{\alpha}-1}\Big[\sum_{j=1}^{n_1} \gamma_j^{\frac{1}{\alpha}} V_j \boldsymbol{s}_j + \sum_{j=n_1+1}^{n_1+m} (2\gamma_j)^{\frac{1}{\alpha}} W_j \boldsymbol{s}_j\Big] + \boldsymbol{\mu}\Big].\end{aligned}$$

2. Case of $\alpha = 1$: From 2, 4, and 7, we have

$$\begin{aligned}\boldsymbol{X} &\stackrel{d}{=} Z\boldsymbol{Y} + (\frac{2}{\pi}Z\log Z)\boldsymbol{g} + \boldsymbol{\mu}Z \\ &\stackrel{d}{=} Z\Big[\sum_{j=1}^{n_1} \gamma_j(V_j + \frac{2}{\pi}\log\gamma_j)\boldsymbol{s}_j + \sum_{j=n_1+1}^{n_1+m} 2\gamma_j W_j \boldsymbol{s}_j\Big] \\ &\quad + \frac{2}{\pi}Z\log(Z)\sum_{j=1}^{n_1}\gamma_j\boldsymbol{s}_j + \frac{2}{\pi}Z\log(Z)\sum_{j=n_1+1}^{n_1+m}\gamma_j\boldsymbol{s}_j + \boldsymbol{\mu}Z \\ &= Z\Big[\sum_{j=1}^{n_1} \gamma_j(V_j + \frac{2}{\pi}\log(\gamma_j Z))\boldsymbol{s}_j + \sum_{j=n_1+1}^{n_1+m} 2\gamma_j(W_j + \frac{\log Z}{\pi})\boldsymbol{s}_j + \boldsymbol{\mu}\Big]\end{aligned}$$

□

We call this method M2. The following algorithm that follows from theorem 3.2, can be use for faster simulating of geometric stable random vectors.

Algorithm 2.

- Generate n_1 independent random variables $V_j, j = 1, \ldots, n_1$ from $S(\alpha, 1)$.
- Generate m independent random variables $W_j, j = n_1+1, \ldots, n_1+m$ from $S(\alpha, 0)$.
- Generate a standard exponential random variable Z, independent of V_j's and W_j's.
- Compute

$$\boldsymbol{X} = Z^{\frac{1}{\alpha}} B\boldsymbol{W} + \boldsymbol{\Delta}$$

where

$$\boldsymbol{W} = (V_1 + I_{\{1\}}(\alpha)\frac{2}{\pi}\log(\gamma_1 Z), \ldots, V_{n_1} + I_{\{1\}}(\alpha)\frac{2}{\pi}\log(\gamma_{n_1} Z),$$
$$W_{n_1+1} + I_{\{1\}}(\alpha)\frac{\log Z}{\pi}, \ldots, W_{n_1+m} + I_{\{1\}}(\alpha)\frac{\log Z}{\pi})',$$

$$B = (\gamma_1^{\frac{1}{\alpha}}\boldsymbol{s}_1, \ldots, \gamma_{n_1}^{\frac{1}{\alpha}}\boldsymbol{s}_{n_1}, (2\gamma_{n_1+1})^{\frac{1}{\alpha}}\boldsymbol{s}_{n_1+1}, \ldots, (2\gamma_{n_1+m})^{\frac{1}{\alpha}}\boldsymbol{s}_{n_1+m}),$$

and $\boldsymbol{\Delta} = (z\mu_1, \ldots, z\mu_d)'$. The following S-PLUS function can be use for generate N pseudo stable random vectors based on M2 method.

Function name: rmvgstab2

```
function(N, alpha, LG, Ga, mu)
{
 # nSym = number of symmetric points in the last nSym columns
 #        of LG, (corresponding Ga should be doubled)
   nrow <- dim(LG)[1]
   symvec <- symindex <- rep(0, nrow)
   for(i in 1:nrow) {
     j <- i + 1; count <- T
     while(count && j <= nrow) {
           if(all(LG[i, ]  == - LG[j, ])) {
               symvec[i] <- i
               symindex[c(i, j)] <- c(i, j)
               count <- F
           }
           else j <- j + 1
     }
   }
   index <- rbind(symindex, symvec)
   newLG <- rbind(LG[index[1, ] == 0, ], LG[index[2, ] != 0, ])
   newGa <- c(Ga[index[1, ] == 0], 2 * Ga[index[2, ] != 0])
   noSym <- sum(index[2, ] != 0)
   LG <- newLG
   Ga <- newGa
   nSym <- noSym
   n <- dim(LG)[1]; d <- dim(LG)[2]
   Z <- rexp(N)
   n1 <- n - nSym; n2 <- ifelse(n1 == 0, n, n1)
   Z1 <- matrix(Z, nrow = N, ncol = d)
   W <- matrix(rstab(N * n2, alpha, sign(n1)) + if(n1 == 0) {
          if(alpha == 1) rep((1/pi) * log(Z), each = n2) else 0 }
          else if(alpha == 1) rep((2/pi) * log(Ga[1:n2]), N) +
                              rep((2/pi) * log(Z), each = n2)
          else tan((pi * alpha)/2), ncol = n2, byrow = T)
   if(nSym > 0 && nSym < n)
   W <- cbind(W, matrix(rstab(N * nSym, alpha) + if(alpha == 1)
              rep((1/pi) * log(Z), each = nSym) else 0,
              ncol = nSym, byrow = T))
   Ga1 <- matrix(Ga, nrow = n, ncol = d)
   B <- LG * Ga1^(1/alpha)
   mu <- matrix(mu, nrow = N, ncol = d, byrow = T) * Z1
   (W %*% B) * Z1^(1/alpha) + mu
}
```

4. simulation study

In order to compare the new method M2 with previous method M1, we run rmvgstab1 and rmvgstab2, using a laptop PC by HP, model G60, equipped with 1.90 GHz AMD Athlon

Dual-Core Ql-60 processor, 2.00 GB of RAM and Windows Vista™ home premium operating system.

We generate samples of size 500,000 from multivariate geometric stable random vectors of M1 and M2 with location of point masses are given in table 1. Table 2 shows the averages of execution times (in seconds) that are based on 100 simulations using each generator.

Table 1. Location of point masses for simulating samples of multivariate geometric stable distribution.

Case	symmetric points	nonsymmetric points
(a)	(1,0)(-1,0)	(0,1)
(b)	(1,0)(-1,0)$(\frac{1}{2},\frac{\sqrt{3}}{2})(\frac{-1}{2},\frac{-\sqrt{3}}{2})$	$(\frac{-1}{2},\frac{\sqrt{3}}{2})(\frac{\sqrt{3}}{2},\frac{-1}{2})$
(c)	(1,0)(-1,0)	$(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})(\frac{-\sqrt{2}}{2},\frac{\sqrt{2}}{2})$
(d)	(1,0)(-1,0)(0,1)(0,-1)	$(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})$
(e)	$(\frac{-1}{2},\frac{\sqrt{3}}{2})(\frac{1}{2},\frac{-\sqrt{3}}{2})$	(1,0)
(f)	(1,0)(-1,0)(0,1)(0,-1)	
(g)	$(\frac{\sqrt{2}}{4},\frac{\sqrt{2}}{4},\frac{\sqrt{3}}{2})(\frac{-\sqrt{2}}{4},\frac{-\sqrt{2}}{4},\frac{-\sqrt{3}}{2})$ $(\frac{-\sqrt{2}}{4},\frac{\sqrt{6}}{4},\frac{\sqrt{2}}{2})(\frac{\sqrt{2}}{4},\frac{-\sqrt{6}}{4},\frac{-\sqrt{2}}{2})$	$(\frac{-1}{2},0,\frac{\sqrt{3}}{2})$
(h)	$(\frac{\sqrt{2}}{2},0,0,\frac{-\sqrt{2}}{2})(\frac{-\sqrt{2}}{2},0,0,\frac{\sqrt{2}}{2})$ $(\frac{-1}{2},\frac{1}{2},\frac{1}{2},\frac{-1}{2})(\frac{1}{2},\frac{-1}{2},\frac{-1}{2},\frac{1}{2})$	$(0,\frac{\sqrt{3}}{3},\frac{-\sqrt{3}}{3},\frac{-\sqrt{3}}{3})$

Table 1. Execution time averages (in seconds) along with standard deviations (in brackets), for simulating samples of size 500,000 from method of M1 and M2 with $\alpha = 0.75, \boldsymbol{\mu} = \mathbf{0}, \gamma_j = 0.25$, and $\gamma_j = 0.5$ for symmetric and nonsymmetric locations, respectively.

Case	M1 method	M2 method
(a)	3.33938(0.1598597)	2.30244(0.1347350)
(b)	4.95785(0.3175116)	4.15462(0.2992480)
(c)	3.85863(0.1990990)	2.89112(0.1814689)
(d)	4.48686(0.2561435)	3.47789(0.2252744)
(e)	3.36542(0.1970318)	2.84639(0.1843554)
(f)	3.99871(0.2237170)	2.29487(0.1512818)
(g)	5.31875(0.2303287)	3.34969(0.1950943)
(h)	6.07894(0.2689873)	3.88684(0.2443736)

References

1. T. J. Kozubowski and A. K. Panorska, *Multivariate geometric stable distributions in financial applications* (Math. Comput. Modelling **29**, 1999), pp. 83–92.
2. R. Modarres and J. P. Nolan, *A method for simulating stable random vectors* (Comput. Statist. **9**, 1994), pp. 11–19.
3. T. J. Kozubowski and A. K. Panorska, *Simulation of geometric stable and other limiting multivariate distributions arising in random summation scheme* (Math. Comput. Modelling **29**, 1999), pp. 255–262.

4. H. Zare, M. Firouzi and A. Mohammadpour, *An improved method for simulation of stable and geometric stable random vectors (in Press)*, (2010).

Statistical Analysis of Power in Cross Classified Models via Simulation

M. Golalizadeh

Department of Statistics Tarbiat Modares University

The multilevel models are used for the application of statistical models to problems where data exhibit an underlying hierarchical structure. They are simply a generalization of both linear regression models and analysis of variance. When the individuals are belonging to more than one group units the suitable models to deal with these situations are the cross classified (XC) models in which the units are cross classified as well as clustered. Determination of sample size in the multilevel models were mostly concentrated on either simple variance components model or empirical studies on exploring the effects of some sampling components. We propose simple simulation algorithm to determine the optimum sample size combinations. It is more complicated in XC models due to complexity of structural design. Two simple approaches are described and then compared in an example using fake data constructed as XC model.

Keywords: Multilevel models, Cross classified models, Design effect, Sample size, Power curve.

1. Introduction

The multilevel models are generalization of classical regression and analysis of variance. The model are applied to data which exhibit a hierarchical structure, such as students within schools, patients within wards, etc (Goldstein, 1995)[[1]] . Implementing these models in diverse fields of sciences including sociology, medicine, geography, politics and so on made a great attention over the last three decades or so. In all these fields, the number of subjects to be considered such that a predefined effect will be statistically significant is of great important to answer some scientific questions. In another word, the researchers in these fields are interested on determining the sample size for the subjects in various levels before performing an observational or experimental study. This issue inherently invokes the power, hypothesis test and intensive computer programming.

It is well known that the cluster sampling can be performed, at least, in two ways (Cochran, 1977)[[2]]. They are one-stage and two-stage cluster sampling. Other popular sampling techniques in this framework are stratified cluster sampling and sampling with probability proportional to size. It sounds that sampling in the multilevel models are mostly similar to cluster sampling and so the procedure to compute the sample size is easy. However, there are some challenges, particularly in XC models.

Most activities in sample size calculation in multilevel models focused on simulation studies with some exceptions. For example, Bosker et al. (2003) have provided a computer programme called PINT in performing power analysis for two level models [[3]]. The

programme uses the approximate standard errors for the parameters in two level models and assume the balance at level two unit. The theoretical issue behind the programme is given in Snijder and Bosker (1993)[[4]]. However, their programme is not based upon simulation study. Hence, we cannot use PINT here and make a comparison with our simulation approach. Other research including Afshartous (1995)[[5]] and Mok (1995)[[6]] consider the empirical studies for particular problem in multilevel model aiming to explore the impact of random effects in the model based upon some combinations of sample sizes through simulation. Our method is, spiritually, similar to the procedure suggested by Gelman and Hill (2007, Ch. 20), except they illustrate the approach for a particular variance components and two level models [[7]]. The procedure in this paper has already been implemented in the free software MLPOWSIM written by the author and his colleagues (Browne, et al. 2009)[[8]]. It is worth to mention that tackling sample size determination in multilevel models from statistical Bayesian view is studied by Wang and Gelfand (2002)[[9]].

In Section 2, we present the sample size determination in variance components model as a motivation. In Section 3, we propose a theoretical procedure to calculate the sample size in XC models. A simulation study for these models based upon particular fake data is given in Section 4. The paper is ended with conclusion.

2. Sample Size Determination in Variance Components Model

To start our discussion, the following approximate equation, for testing equality of the population mean of a univariate Normal distributed variable to a fixed value, illustrates the relationship between the sample size and other sampling parameters (see, for example, Bain and Engelhardt 2000 [[10]]).

$$\frac{\gamma}{s.e.(\gamma)} \simeq Z_{1-\alpha} + Z_{1-\beta}, \tag{1}$$

where $s.e.(\gamma)$ denotes the standard error of γ (or of the parameter of interest, if we want to be precise in statistical sense) and Z_q is q-quantile of the standard Normal distribution. Note that, assuming the sampling fraction is approximately one, the value of the sample size (n) appearers in above equation as $s.e.(\gamma) = s.d.(\gamma)/\sqrt{n}$, where $s.d.(\gamma)$ stands for the standard deviation of γ.

In multilevel models, one cannot directly use the equation (1) to derive the optimum sample sizes. Instead, we describe a simple method to derive approximate power and choose proper sample size using simulation algorithm. To clarify the subject, we describe the procedure for the variance components model. Method can be easily extended to other multilevel models.

The variance components model is defined as (see, for example, Goldstein, 1995 [[1]])

$$y_{i\,j} = \beta_0 + u_j + \epsilon_{i\,j}, \qquad j = 1, \ldots, N, \quad i = 1, \ldots, n \tag{2}$$

where $u_j \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_u^2)$ and $\epsilon_{i\,j} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$, and the random residuals u_j and $\epsilon_{i\,j}$ are assumed to be independent. Note that the model is balanced at level two, i.e. equal numbers of level 1 units per level two. To simulate hierarchical data from the model (2), we can write a function for various combination of the lower (n) and higher (N) level units while the other parameters, i.e. β_0, σ_u^2, and σ_e^2, are fixed in advance (to be given by the user in his/her preference). Then, the power for each sample combination can be

calculated using any statistical techniques. We present a simple method here based upon converting the confidence interval.

Assume the interest is to test the hypothesis

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}$$

at the significant level α. Since the user knows the sign (positive or negative) of the fixed effects in advance, a fair decision can be made in rejecting the null hypotheses based upon the upper and lower limits for particular sample size combination. In this particular case (and also in other circumstances), if the lower (upper) bound is greater (less) than zero, the null hypothesis can be rejected. Now, the user can simulate the model given in (2) for the large number of times, say 1000, and approximate the power of test to be the percentage of times when the null hypothesis is rejected. This procedure can be repeated for other sample size combinations. A simple inspection through a table or plot give an impression for the required sample size to attain an specific power.

From theoretical point of view, the design effect can be utilized to derive the power for particular sampling design. This, then, is used to check whether or not the approximated power given by simulation studies based upon sample size combination under a new design has attained the power derived from (1). To justify this, let us consider the balanced model given by (2), with n lower units in each of N upper units. Then, the design effect ($DEFF$) of this model is (see, for example, Snijder and Bosker, 1993 [[4]])

$$DEFF = 1 + (n-1)\rho,$$

where ρ is the intraclass correlation. Now, if we know the required sample size for a given power in a simple random sample (*srs*) design we should multiply it by the above design effect to get corresponding sample in the model given by (2).

Now, we derive the design effect for the XC models before using it in a simulation study.

3. Cross Classified Models

The implemented notations here are the same as those described in Goldstein (1995) [[1]]. Consider the null XC model

$$y_{i\,(\,j\,k)} = \beta_0 + u_j^{(2)} + v_k^{(3)} + e_{i\,(\,j\,k)} \quad i = 1, \ldots, n,\ j = 1, \ldots, n_u\,, k = 1, \ldots, n_v, \qquad (3)$$

where $u_j^{(2)} \sim N(0, \sigma_u^2)$, $v_k^{(3)} \sim N(0, \sigma_k^2)$ and $e_{i\,(\,j\,k)} \sim N(0, \sigma^2)$ and a further assumption in which these three errors are mutually independents. It can be seen that this model is a balance model in which the number of observations for each combination of the first and second XC factors are fixed n in each combination of the first and second XC samples, i.e. n_u and n_v, respectively.

The estimator of β in (3), with superscript xc referring to XC model, is

$$\widehat{\beta_0^{\mathrm{xc}}} = \frac{\sum_{i\,j\,k} y_{i\,(\,j\,k)}}{n\,n_u\,n_v},$$

and its variance is

$$\mathrm{Var}\big(\widehat{\beta_0^{\mathrm{xc}}}\big) = \frac{\mathrm{Var}\Big(\sum_{i\,j\,k} y_{i\,(\,j\,k)}\Big)}{(n\,n_u\,n_v)^2}. \qquad (4)$$

Clearly, the expression (4) involves the variance of $y_{i\,(\,j\,k)}$ and covariance of distinct $y_{i\,(\,j\,k)}$'s. Using (3), the variance term turns out as $\sigma_u^2+\sigma_v^2+\sigma^2$. However, outcome of the covariance terms depend on the how the indexes i, j and k appear in distinct $y_{i\,(\,j\,k)}$'s. It is seen from (3) that, for every i, j, and k,

$$\begin{aligned}
&\mathrm{Cov}(y_{i\,(\,j\,k)}, y_{i'\,(\,j\,k)}) = \sigma_u^2+\sigma_v^2, \qquad i\neq i',\\
&\mathrm{Cov}(y_{i\,(\,j\,k)}, y_{i'\,(\,j'\,k)}) = \sigma_v^2, \qquad i\neq i', \quad j\neq j',\\
&\mathrm{Cov}(y_{i\,(\,j\,k)}, y_{i'\,(\,j\,k')}) = \sigma_u^2, \qquad i\neq i', \quad k\neq k'.
\end{aligned} \tag{5}$$

Now, using these information we can expand the expression (4) as follows:

$$\begin{aligned}
\mathrm{Var}\big(\widehat{\beta_0^{\mathrm{xc}}}\big) &= \frac{1}{(n\,n_u\,n_v)^2}\Big[nn_un_v(\sigma^2+\sigma_u^2+\sigma_v^2)+n(n-1)n_un_v(\sigma_u^2+\sigma_v^2)\\
&\quad +n^2n_un_v(n_v-1)\sigma_u^2+n^2n_vn_u(n_u-1)\sigma_v^2\Big]\\
&= \frac{\sigma^2+\big[1+n-1+n(n_v-1)\big]\sigma_u^2+\big[1+n-1+n(n_u-1)\big]\sigma_v^2}{n\,n_u\,n_v}\\
&= \frac{\sigma^2+n\,n_v\sigma_u^2+n\,n_u\sigma_v^2}{n\,n_u\,n_v}.
\end{aligned} \tag{6}$$

In order to compare the efficiency of XC modeling with *srs* allocation of samples, we do need to derive the variance of the estimator for the parameter of the model, i.e. $\widehat{\beta}_0$, under the *srs* design with the same numbers of observation, i.e. $n\,n_u\,n_v$. Since under the *srs* design the random variables $y_{i\,(\,j\,k)}$'s given in (3) are independent samples from the Normal distribution with mean zero and variance $\sigma^2+\sigma_u^2+\sigma_v^2$, we can expand (4) for *srs* design as

$$\begin{aligned}
\mathrm{Var}\big(\widehat{\beta_0^{srs}}\big) &= \frac{var\Big(\sum_{i\,j\,k} y_{i\,(\,j\,k)}\Big)}{(n\,n_u\,n_v)^2} = \frac{\sum_{i\,j\,k}\,var(y_{i\,(\,j\,k)})}{(n\,n_u\,n_v)^2}\\
&= \frac{\sum_{i\,j\,k}(\sigma^2+\sigma_u^2+\sigma_v^2)}{(n\,n_u\,n_v)^2} = \frac{\sigma^2+\sigma_u^2+\sigma_v^2}{n\,n_u\,n_v}.
\end{aligned} \tag{7}$$

Consequently, the relative efficiency of the *srs* design with respect to the XC design is the ratio of the variances in (6) to (7). It is the *DEFF* of the XC design in (3), simplified as,

$$DEFF = \frac{\sigma^2+n\,n_v\sigma_u^2+n\,n_u\sigma_v^2}{\sigma^2+\sigma_u^2+\sigma_v^2}. \tag{8}$$

If we define the intraclass correlation coefficients of the first and second XC factors, respectively, as $\rho_{I_u}=\frac{\sigma_u^2}{\sigma^2+\sigma_u^2+\sigma_v^2}$ and $\rho_{I_v}=\frac{\sigma_v^2}{\sigma^2+\sigma_u^2+\sigma_v^2}$, then (8) can be written as

$$\mathrm{DEFF} = 1+(n\,n_v-1)\rho_{I_u}+(n\,n_u-1)\rho_{I_v}. \tag{9}$$

As a result, if we know the total sample size and further assume the effect size, variances and the significant level are given in advance we could derive the power of test about β_0 with multiplying the above design effect in the sample size obtained with *srs* design and, then, plugging it in the formula (1).
Next, recalling the above discussion we describe a simulation study and present the results in following section.

4. A Simulation Study

We are encountered with two simulation scenarios here. One is that dealing with simulating the model under study (to generate fake data) and the another concerns about how the power analysis should be carried out. To deal with the first scenario, one can benefit the facility of the statistical softwares such as `MLwiN`, `SAS` and others. We, instead, write a programme in `R` mimicking the XC structure represented by (3). The estimate of model parameters given by the function `lmer` from the package `lme4` in `R` have been further compared with popular multilevel software `MLwiN`. The results were quite similar, although the latter software are faster than the former.

To deal with the second scenario, i.e. obtaining the power of the fixed parameter, we have implemented two approaches. Obviously, there might be some other approaches in literature. However, our approaches are very easy to implement in programming and also have interesting intuitive interpretation.

The first approach is the same procedure as described for variance components model. That is to evaluate, approximately, whether or not the effect size is inside our estimated confidence interval. In a long simulation, say *Iteration*, we consider the average of 0/1 values as an approximate value for the power. We name this approach as 0/1 approach and indicate the power obtained by this approach with $z^{0/1}$. We, further, derive an approximate confidence interval for power in this approach with the lower and upper limits, respectively, $z_L^{0/1}$ and $z_U^{0/1}$. Since the approximate power follows a Binomial probability function with the number of trials *Iteration* and probability of success $z^{0/1}$, using the Normal approximation to Binomial distribution, the $100(1-\alpha)\%$ confidence interval for the approximate power turns out as

$$(z^{0/1} - Z_{\frac{\alpha}{2}}\sqrt{\frac{z^{0/1}(1-z^{0/1})}{Iteration}}\,,\, z^{0/1} + Z_{\frac{\alpha}{2}}\sqrt{\frac{z^{0/1}(1-z^{0/1})}{Iteration}}).$$

In the second approach, we indirectly employ the approximate equality in (1). The procedure is as follows:

Assume in the first iteration of our simulation, i.e. *Iteration=1*, we fit the model given by (3) for our simulated data and get the standard error $s.e.(\beta_0)$. This can be added to the standard error obtained in the second stage (*Iteration=2*) of model fitting and so on. At the end, the average of these standard errors can be considered as an estimate for the standard error $s.e.(\beta_0)$. Now, having the standard error and effect size of β_0 along with the significant level α, we can utilize (1) to derive an estimate for the power. In addition, we can obtain the empirical variance, say $\mathrm{Var}(z^s)$, of the standard errors given in each iteration to build a confidence interval for the power. Calling this method as s approach, we name the power derived in this approach by z^s and the corresponding confidence bounds by z_L^s and z_U^s. Note that the $100(1-\alpha)\%$ confidence interval for the power in this approach will be

$$(z^s - Z_{\frac{\alpha}{2}}\sqrt{\frac{\mathrm{Var}(z^s)}{Iteration}}\,,\, z^s + Z_{\frac{\alpha}{2}}\sqrt{\frac{\mathrm{Var}(z^s)}{Iteration}}\,).$$

We have programmed these approaches along with constructing the model (3) in `R`. Some sampling parameters have been fixed as follows:

$$n = 5,\ n_u = n_v = 20,\ \alpha = 0.05,\ Iteraion = 1000.$$

We further fix the value $\sigma^2+\sigma_u^2+\sigma_v^2$ to 10 and the effect size to 0.1728 such that, using these values in the *srs* design and equation (1), the power of test about β_0 turns out as $1-\beta \simeq 0.79$.

Keeping the value $\sigma^2+\sigma_u^2+\sigma_v^2$ fixed to 10, we alter variances in our simulation study while fitting XC model represented by (3). Then, power of test about β_0 using both two approaches along with their corresponding lower and upper limits were obtained. These values along with different values of variance parameters are given in Table 1. Note that the theoretical power value in each row can be obtained by solving a modification of (1). For example, using (9) the design effect for the fifth row with parameter values $\sigma_u^2=\sigma_v^2=0.5$, and $\sigma^2=9$, is 10.9. So, we have

$$Z_{1-\beta}=\frac{0.1728}{\sqrt{\frac{10}{2000}DEFF}}-1.645=\frac{0.1728}{\sqrt{\frac{10}{2000}10.9}}-1.645=-0.907.$$

Consequently, the theoretical power for the fifth row will be $1-\beta=0.18$. Similarly, for the other rows the theoretical power can be derived. They are 0.79, 0.78, 0.72, 0.41, from the first to fourth row, respectively. Having these value, one can check which approach is more sensible compared with another. It is seen that as the variance of measurement error decreases there is less power to reject $\beta_0=0$. The length of CI using 0/1 approach is about 5% overally while it is about 1% in s approach, indicating the accuracy of the latter approach. Moreover, for the higher variance of measurement error, the CI using s approach cover the true value of the power more precisely.

Table 1. Approximate powers and their 95% CIs for β_0 using 0/1 and s approaches

Variances in XC model (3)		$z_L^{0/1}$	$z^{0/1}$	$z_U^{0/1}$	z_L^s	z^s	z_U^s
$\sigma_u^2=\sigma_v^2=0$	$\sigma^2=10$	0.722	0.749	0.776	0.705	0.710	0.715
$\sigma_u^2=\sigma_v^2=0.001$	$\sigma^2=9.998$	0.706	0.733	0.760	0.699	0.705	0.710
$\sigma_u^2=\sigma_v^2=0.01$	$\sigma^2=9.98$	0.675	0.703	0.731	0.668	0.674	0.680
$\sigma_u^2=\sigma_v^2=0.1$	$\sigma^2=9.8$	0.412	0.443	0.474	0.409	0.414	0.419
$\sigma_u^2=\sigma_v^2=0.5$	$\sigma^2=9$	0.174	0.199	0.224	0.183	0.184	0.186

5. Conclusion

The traditional approaches to determine the sample size for multilevel surveys is more involved than simple design structure. With availability of fast computing softwares, the simulation based study to perform power analysis is paid more attention. We presented two simple methods to study the sample size determination in multilevel cross classified models using simulation studies. As the approaches are simple and intuitively tractable, it can be, roughly speaking, implemented for other multilevel models. It is more user friendly than traditional method in which one needs to know the sample before carrying out a research to investigate some scientific questions.

Acknowledgement

I would like to thank William Browne on pointing out the problem and many useful comments.

References

1. H. Goldstein, *Multilevel Statistical Models*, (Edward Arnold, London, 1995).
2. W. G. Cochran, *Sampling Techniques*, 3rd edition (Wiley, New York, 1977).
3. R. J. Bosker, T. A. B. Snijders and H. Guldemond, *PINT (Power IN Two-level designs) User Manual* (Groningen/Enschede, 2003)
4. T. A. B. Snijders and R. J. Bosker, Standard errors and sample size for two-level research, *Journal of Educational Statistics* **18**, 237-259 (1993).
5. D. Afshartous, Determination of Sample Sizes for Multilevel Model Design, *Paper presented at the annual meeting of the American Educational Research Association*, San Francisco, 18-22 (1995).
6. M. Mok, Sample Size Requirements for 2-level Designs in Educational Research, *Multilevel Modelling Newsletter*, **7**, 11-15 (1995).
7. A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, (Cambridge University Press, New York, 2007)
8. W. J. Browne, M. Golalizadeh, R. MA. and Parker, A Guide to Sample Size Calculations for Random Effect Models via Simulation and the MLPowSim Software Package, *Bristol University* (2009).
9. F. Wang, and A. E. Gelfand, A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models, *Statistical Science*, **17**,193-208 (2002).
10. L. J. Bain, and M. Engelhardt, *Introduction to Probability and Mathematics Statistics*, (Duxbury, 2000).

Using Approximate MLE for Testing Normality Based on Kullback-Leibler Information with Progressively Type-II Censored Data

A. Habibi Rad and F. Yousefzadeh

Department of Statistics, Ferdowsi University of Mashhad

Department of Statistics, Birjand University

We will use the joint entropy of progressively censored order statistics in terms of an incomplete integral of the hazard function, and provide a simple estimate of the joint entropy of progressively Type-II censored data, has been introduced by Balakrishnan et al. (2007). Then We construct a goodness-of-fit test statistic based on Kullback-Leibler information for Normal distribution by using approximate MLE. Finally, we used Monte Carlo simulations, the power of the test is estimated and compared against several alternatives under different progressive censoring schemes.

Keywords: Approximate Maximum Likelihood Estimate, Entropy, Goodness-of-fit test, Hazard function, Monte Carlo simulation, Progressively Type-II censored data.

1. Introduction

Suppose a random variable X has a distribution function $F(x)$ and a continuous density function $f(x)$. The differential entropy $H(f)$ of the random variable is defined in Shannon (1948), to be

$$H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx. \tag{1}$$

The first time, the test normality performed based on sample entropy by Vasicek (1976) and the power compared with some leading test statistics for complete samples.
The entropy difference $H(f) - H(g)$ has been considered in Dudewicz et al. (1981) and Gokhale (1983) for establishing goodness-of-fit tests for the class of the maximum entropy distributions.

The Kullback-Leibler (KL) information in favor of $g(x)$ against $f(x)$ is defined in Kullback (1959) to be

$$I(g : f) = \int_{-\infty}^{\infty} g(x) \log \frac{g(x)}{f(x)} dx,$$

which is an extended concept of entropy.
Because $I(g : f)$ has the property that $I(g : f) \geq 0$, and the equality holds if $g = f$, the estimate of the KL information has also been considered as a goodness-of-fit test statistic by some authors including Arizono et al. (1989) and Ebrahimi et al. (1992), for complete

samples. Park (2005) and Balakrishnan et al. (2007), respectively, for Type-II censored data and progressively Type-II censored data.
Now, in this paper we will extend the goodness-of-fit test based on KL information with progressively Type-II censored data for Normal distribution.

The rest of the paper is organized as follows: In Section 2 as Preliminary, we introduce Type-II progressive censoring data, the joint entropy of progressively censored data in terms of the hazard function and the nonparametric estimate of the joint entropy. In Section 3, we define the KL information for progressively Type-II censored data and propose a goodness-of-fit test for Normality based on KL information, in Section 4. Finally, in Section 5 we use Monte Carlo simulations to evaluate the power under different Type-II progressive censoring schemes.

2. PRELIMINARY

2.1. *Progressively Type-II Censored Data*

Suppose n identical items are placed on a life-testing experiment. Assume that their lifetimes are independent and identically distributed with probability distribution function (cdf) $F(x;\underline{\theta})$ and probability density function (pdf) $f(x;\underline{\theta})$, where θ is a vector of parameters.

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 unites. The two most common censoring schemes are termed as conventional Type-I and Type-II censoring schemes which are extensively studied in statistical and reliability literature, Balakrishnan and Cohen (1991). Briefly, they can be described as follows: Consider n items under observations in a particular experiment. In the conventional Type-I censoring scheme, the experiment continues up a pre-specified time T. The conventional Type-II censoring scheme requires the experiment to continue until a pre-specified number of failures $m(\leq n)$ occur.
One of the drawbacks of the conventional Type-I, Type-II censoring schemes is that they do not allow for removal of units at points other than the terminal point of the experiment. One censoring scheme known as Type-II progressive censoring scheme, which has this advantage, so it becomes very popular for the last few years. It can be described as follows: Consider n units in a study and suppose $m(\leq n)$ is fixed before the experiment. Moreover, m other integers, $R_1, \cdots R_m$ are also fixed before so that $R_1+\cdots+R_m+m=n$. At the time of the first failure, say $X_{1:m:n},\ R_1$ of the remaining units are randomly removed. Similarly, at the time of the second failure, say $X_{2:m:n},\ R_2$ of the remaining units are randomly removed and so on. Finally, at the time of the $m-th$ failure, say $X_{m:m:n}$, the rest of the R_m units are removed. Fore further details on Type-II progressive censoring, refer to Balakrishnan and Aggarwala (2000).

The joint probability density function (pdf) of all m progressively Type-II censored order statistics $(X_{1:m:n}, \cdots X_{m:m:n})$ which is define in Balakrishnan (2000) to be

$$f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1,x_2,\cdots,x_m) = c\prod_{i=1}^{m} f(x_i)\{1-F(x_i)\}^{R_i},$$

$$x_1 < x_2 < \cdots < x_m,$$

where

$$c = n(n - R_1 - 1)\cdots(n - R_1 - R_2 - \cdots - R_{m-1} - m + 1).$$

2.2. *Entropy of Progressively Censored Data in Terms of the Hazard Function*

The joint entropy of $X_{1:m:n}, \cdots, X_{m:m:n}$ defined in literature (Park, 2005), to be

$$H_{1\cdots m:m:n} = -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{x_{2:m:n}} f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m) \times \log f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m)dx_{1:m:n}\cdots dx_{m:m:n},$$

where $f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m)$ is the joint pdf of all m progressively Type-II censored order statistics.

$H_{1\cdots m:m:n}$ is an m-dimensional integral, and we need to simplify this multiple integral.

The simple calculation of the entropy of the usual single and consecutive order statistics has been studied in Wong et al. (1990) and Park (1995). The multiple integral of the entopy for Type-II censored data be simplified to a single-integral by Park (2005) and the joint entropy of progressively Type-II censored order statistics in terms of an incomplete integral of the hazard function , $h(x)$, has been simplified by Balakrishnan et al. (2007),

$$H_{1\cdots m:m:n} = -\log c + n\overline{H}_{1\cdots m:m:n},$$

where

$$\overline{H}_{1\cdots m:m:n} = \frac{m}{n} - \frac{1}{n}\int_{-\infty}^{\infty}\sum_{i=1}^{m} f_{X_{i:m:n}}(x)\log h(x)dx.$$

2.3. *Nonparametric Entropy Estimate*

The nonparametric estimate of the joint entropy ($H_{1\cdots m:m:n}$) was obtained, as

$$H_{1\cdots m:m:n}(w, n, m) = -\log c + nH(w, n, m),$$

where

$$H(w, n, m) = \frac{1}{n}\sum_{i=1}^{m}\log\left(\frac{(x_{i+w:m:n} - x_{i-w:m:n})}{E(U_{i+w:m:n}) - E(U_{i-w:m:n})}\right) - (1 - \frac{m}{n})\log(1 - \frac{m}{n}).$$

(Balakrishnan et al., 2007).

3. Goodness-of-fit Test Based on the Kullback-Leibler Information

For a null density function $f^0(x;\theta)$, the KL information from a progressively Type-II censored data is given by

$$I_{1\cdots m:m:n}(f : f^0) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{x_{2:m:n}} f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m;\theta) \times \log\frac{f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m;\theta)}{f^0_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m;\theta)}dx_1\cdots dx_m,$$

where $f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m)$ is the joint pdf of all m progressively Type-II censored order statistics.

The KL information can be estimated by

$$I_{1\cdots m:m:n}(f : f^0) = -n\overline{H}_{1\cdots m:m:n} - \sum_{i=1}^{m} \log f^0(x_i;\theta) - \sum_{i=1}^{m} R_i \log\left(1 - F^0(x_i;\theta)\right). \quad (2)$$

Thus, the test statistic based on $\frac{1}{n} I_{1\cdots m:m:n}(f : f^0)$ is given by

$$T(w,n,m) = -H(w,n,m) - \frac{1}{n}\left[\sum_{i=1}^{m} \log f^0(x_i;\widehat{\theta}) + \sum_{i=1}^{m} R_i \log(1 - F^0(x_i;\widehat{\theta}))\right], \quad (3)$$

where $\widehat{\theta}$ is an estimation of θ.

4. Test for Normality

Suppose we are interested in goodness-of-fit test for
$H_0 : f^0 = (2\pi\sigma^2)^{\frac{-1}{2}} exp\{-(x-\mu)^2/2\sigma^2\}$ vs $H_A : f^0 \neq (2\pi\sigma^2)^{\frac{-1}{2}} exp\{-(x-\mu)^2/2\sigma^2\}$
where $\underline{\theta} = (\mu, \sigma^2)$ is unknown.
Then, the KL information for a progressively Type-II censored data can be approximated, by (3) and we estimate the unknown parameters (μ, σ^2) by the maximum likelihood estimate (MLE).
The MLE for progressively Type-II censored sample from a $Normal(\mu, \sigma^2)$ distribution obtain by solving the below equations, (Balakrishnan and Aggarwala, 2000)

$$\frac{\sum_{i=1}^{m} x_i}{m} = \overline{x} = \mu - \frac{\sigma}{m}\sum_{i=1}^{m} R_i Z_i,$$

$$\frac{\sum_{i=1}^{m}(x_i - \overline{x})^2}{m} = s^2 = \sigma^2\{1 - \frac{1}{m}\sum_{i=1}^{m} R_i \xi_i Z_i - (\frac{1}{m})^2 \sum_{i=1}^{m}(R_i Zi)^2\},$$

where $Z_i = \frac{\varphi(\xi_i)}{1-\phi(\xi_i)}$ and $\varphi(.)$ is the probability density function of the standard normal distribution.
At the first we used a simple iterative procedure such as Newton's method for solving the above equations, but the MLE can not be obtained in explicit form so the next section we propose the approximate maximum likelihood estimates which have explicit forms.

4.1. *Approximate Maximum Likelihood Estimates for Normal Distribution*

In this section, we use the approximate maximum likelihood estimation method (AMLE) developed by Balakrishnan (1989 a,b, 1990 a,b,c) to estimate the scale and location parameters μ and σ.The likelihood function based on progressive Type-II censored sample $x_{1:m:n}, ..., x_{m:m:n}$ with censoring scheme $R_1, ..., R_m$ can be written as

$$L(\mu,\sigma) = c\frac{1}{\sigma^m}\prod_{i=1}^{m} f(z_{i:m:n})(\bar{F}(z_{i:m:n}))^{R_i},$$

where $c = n(n-R_1-1)\cdots(n-R_1-R_2-\cdots-R_{m-1}-m+1)$, $z_{i:m:n} = \frac{x_{i:m:n}-\mu}{\sigma}$, $\overline{F}(.) = 1-F(.)$ and f, F are the probability density function(pdf) and cumulative distribution function (cdf) of Normal standard distribution, respectively.
Upon partial differentiation of the logarithm of the likelihood function with respect to μ and σ, the score equations to be solved for μ and σ in this case are given by

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{m} z_{i:m:n} + \frac{1}{\sigma}\sum_{i=1}^{m} R_i \frac{f(z_{i:m:n})}{\bar{F}(z_{i:m:n})} = 0 \tag{4}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{m}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{m} z_{i:m:n}^2 + \frac{1}{\sigma}\sum_{i=1}^{m} R_i z_{i:m:n} \frac{f(z_{i:m:n})}{\bar{F}(z_{i:m:n})} = 0, \tag{5}$$

Clearly, (4) and (5) do not have explicit solutions. We expand the function $\frac{f(z_{i:m:n})}{\bar{F}(z_{i:m:n})}$ in Taylor series around the point $\xi_i = F^{-1}(p_i)$, where $p_i = 1-q_i = 1-\prod_{j=m-i+1}^{m}\alpha_j$. Balakrishnan and Aggarwala(2000) deduced that: if $U_{i:m:n}, \quad i = 1,\cdots,m$ denote a progressive Type-II censored sample from the uniform(0, 1) distribution obtained from a sample of size n with the censoring scheme $(R_1,\cdots,R_m)$, then $V_i, i = 1,\cdots,m$ are all independent random variables with $V_i = Beta(i + \sum_{j=m-i+1}^{m} R_j, 1),\ i = 1,\cdots,m$, such that

$$U_{i:m:n} = 1 - \prod_{j=m-i+1}^{m} V_j, \quad i = 1,\cdots,m,$$

and

$$E(U_{i:m:n}) = 1 - \prod_{j=m-i+1}^{m} \alpha_j, \quad i = 1,\cdots,m,$$

where

$$\alpha_j = \frac{j + \sum_{i=m-j+1}^{m} R_i}{1 + j + \sum_{i=m-j+1}^{m} R_i}, \quad j = 1,\cdots,m.$$

Then we consider the following approximations

$$\frac{f(z_{i:m:n})}{\bar{F}(z_{i:m:n})} \simeq \alpha_i + \beta_i z_{i:m:n}, \tag{6}$$

where

$$\alpha_i = \frac{f(\xi_i)}{\bar{F}(\xi_i)} - \xi_i\left[-\xi_i\frac{f(\xi_i)}{\bar{F}(\xi_i)} + \left(\frac{f(\xi_i)}{\bar{F}(\xi_i)}\right)^2\right],$$

$$\beta_i = -\xi_i\frac{f(\xi_i)}{\bar{F}(\xi_i)} + \left(\frac{f(\xi_i)}{\bar{F}(\xi_i)}\right)^2.$$

Using the approximations (6) in (4) and (5), we obtain

$$\sum_{i=1}^{m} z_{i:m:n} + \sum_{i=1}^{m} R_i(\alpha_i + \beta_i z_{i:m:n}) = 0, \tag{7}$$

$$-m + \sum_{i=1}^{m} z_{i:m:n}^2 + \sum_{i=1}^{m} R_i z_{i:m:n}(\alpha_i + \beta_i z_{i:m:n}) = 0. \tag{8}$$

From (7) we obtain the AMLE of μ as

$$\hat{\mu} = B + \hat{\sigma}C,$$

where

$$B = \frac{m\bar{x} + \sum_{i=1}^{m} R_i\beta_i x_{i:m:n}}{m + \sum_{i=1}^{m} R_i\beta_i},$$

$$C = \frac{\sum_{i=1}^{m} R_i\alpha_i}{m + \sum_{i=1}^{m} R_i\beta_i}.$$

From (8), we obtain $\hat{\sigma}$ as a solution of the quadratic equation

$$A_1\sigma^2 + A_2\sigma + A_3 = 0,$$

where

$$A_1 = -m, A_2 = \sum_{i=1}^{m} R_i\alpha_i(x_{i:m:n} - B),$$

$$A_3 = \sum_{i=1}^{m}(1 + R_i\beta_i)(x_{i:m:n} - B)^2 > 0.$$

Therefore

$$\hat{\sigma} = \frac{-A_2 - \sqrt{A_2^2 - 4A_1A_3}}{2A_1},$$

is the only positive root.

5. Implementation of Test

Because the sampling distribution of $T(w, n, m)$ is intractable, we determine the percentage points using 10,000 Monte Carlo simulations from Normal distribution. In determining the window size w which depends on n, m and α, we define the optimal window size w to be one which gives minimum critical points. However, we find from the simulated percentage points that the optimal window size w varies much according to m rather than n, and does not vary much according to α, if $\alpha \leq 0.1$. In view of these observations, our recommended values of w for different m are as given in Ebrahimi (1992) and Park (2005).

To obtain the critical values, after deciding about the value of w, simulate the whole procedure by taking the observation from $Normal(0,1)$ distribution and calculate the value of $T(w,n,m)$, for about 10,000 times. Critical values can then be the percentage points of the thus derived (empirical) distribution of T.

5.1. *Power Results for Normal Distribution*

As the proposed test statistic is related to the hazard function of the distribution, we consider the alternatives according to the type of hazard function as follows:

a) Monotone increasing hazard: Gamma and Weibull (shape parameter 2),
b) Monotone decreasing hazard: Gamma and Weibull (shape parameter 0.5),
c) Nonmonotone hazard: Center Beta (shape parameter 0.5),
Log-normal (shape parameter 1).

We used 10,000 Monte Carlo simulations for $n = 10, 20$, to estimate the power of our proposed test statistic. The simulation results are summarized in Tables 1 and 2.

We can see from Tables 1 and 2 that the scheme $(R_1 = 0, \cdots, R_{m-1} = 0, R_m = n-m)$ (the conventional Type-II censored data) shows better power than the other schemes when the alternative is monotone increasing hazard function. For the alternative with monotone decreasing hazard functions, the scheme $(R_1 = n-m, R_2 = 0, \cdots, R_m = 0)$ shows better power; finally, for the alternative with nonmonotone hazard function, sometimes the former censoring scheme gives higher power and sometimes the latter censoring scheme does.

Table 1: Power for different hazard alternatives at 10% significance level for several progressively censored samples when the sample size is $n = 10$.

		monotone increasing hazard alternatives		monotone decreasing hazard alternatives		nonmonotone hazard alternatives	
m	schemes $(R_1, \cdots, R_m)$	Gamma shape 2	Weibull shape 2	Gamma shape 0.5	Weibull shape 0.5	Center Beta shape 0.5	Log-normal shape 1
5	5,0,0,0,0	.198	.123	.555	.705	.341	.403
5	0,5,0,0,0	.196	.118	.563	.700	.387	.388
5	1,1,1,1,1	.165	.116	.516	.624	.417	.274
5	0,0,0,5,0	.123	.101	.435	.566	.312	.214
5	0,0,0,0,5	.173	.133	.464	.539	.422	.235
7	3,0,0,0,0,0,0	.272	.137	.709	.853	.420	.564
7	0,3,0,0,0,0,0	.272	.142	.711	.851	.426	.563
7	1,0,0,1,0,0,1	.262	.148	.704	.825	.507	.486
7	0,0,0,0,0,3,0	.165	.118	.567	.719	.395	.326
7	0,0,0,0,0,0,3	.262	.163	.681	.789	.561	.433
9	1,0,0,...,0,0,0	.319	.147	.808	.926	.501	.660
9	0,1,0,...,0,0,0	.325	.149	.809	.928	.505	.673
9	0,0,...,1,...,0,0	.288	.131	.798	.920	.490	.636
9	0,0,0,...,0,1,0	.249	.141	.735	.876	.506	.542
9	0,0,0,...,0,0,1	.355	.188	.834	.930	.596	.654

Table 2: Power for different hazard alternatives at 10% significance level for several progressively censored samples when the sample size is $n = 20$.

		monotone increasing hazard alternatives		monotone decreasing hazard alternatives		nonmonotone hazard alternatives	
m	schemes $(R_1, \cdots, R_m)$	Gamma shape 2	Weibull shape 2	Gamma shape 0.5	Weibull shape 0.5	Center Beta shape 0.5	LogNormal shape 1
5	15,0,0,0,0	.194	.114	.586	.727	.388	.401
5	0,15,0,0,0	.230	.136	.639	.762	.482	.429
5	3,3,3,3,3	.158	.121	.576	.644	.546	.221
5	0,0,0,15,0	.129	.102	.584	.671	.499	.223
5	0,0,0,0,15	.167	.144	.493	.535	.482	.196
10	10,0,0,...,0,0,0	.354	.148	.906	.973	.667	.745
10	0,10,0,...,0,0,0	.409	.181	.918	.977	.708	.769
10	1,1,1,...,1,1,1	.332	.175	.912	.960	.830	.595
10	0,0,0,...,0,10,0	.146	.123	.651	.788	.507	.265
10	0,0,0,...,0,0,10	.346	.224	.890	.930	.847	.515
15	5,0,0,...,0,0,0	.438	.182	.967	.995	.836	.856
15	0,5,0,...,0,0,0	.472	.204	.969	.996	.846	.869
15	1,1,...,1,...,1,1	.526	.263	.981	.996	.907	.851
15	0,0,0,...,0,5,0	.260	.182	.789	.918	.657	.468
15	0,0,0,...,0,0,5	.558	.302	.982	.996	.934	.833
18	2,0,0,...,0,0,0	.480	.209	.980	.998	.897	.904
18	0,2,0,...,0,0,0	.492	.210	.981	.998	.898	.902
18	1,0,0,...,0,0,1	.630	.287	.996	.999	.948	.940
18	0,0,0,...,0,2,0	.379	.212	.935	.987	.826	.740
18	0,0,0,...,0,0,2	.670	.324	.997	.999	.962	.945

References

Arizono, I. and Ohta, H. " A test for normality based on Kullback-Leibler information," *The American Statistician,* **43**, 20–23, (1989).

Balakrishnan, N. "Approximate MLE of the scale parameter of the Rayleigh distribution with censoring," *IEEE Trans. on Rel.,* **38**, 355–357, (1989a).

Balakrishnan, N. "Approximate maximum likelihood estimation of the mean and standard deviation of the normal distribution based on Type-II censored samples," *J. of Statistical computation and simulation,* **32**, 137-148, (1989b).

Balakrishnan, N. *"Maximum likelihood estimation based on complete and Type-II cencored samples, in the Logistic Distribution,"* (ed. N. Balakrishnan), Marcel Dekker, New York, (1990a).

Balakrishnan, N. "Approximate maximum likelihood estimation for a generalized Logistic distribution," *J. of Statistical Planning and Inference,* **26**, 221-236, (1990b).

Balakrishnan, N. "On the maximum likelihood estimation of the location and scale parameters of exponential distribution based on multiply Type-II censored samples," *J. App. Statist.,* **17**, 55-61, (1990c).

Balakrishnan, N. and Cohen, A. C. *"Order Statistics and Inference: Estimation Methods,"* Academic Press, San Diego, (1991).

Balakrishnan, N. and Aggarwala, R. *"Progressive Censoring: Theory, Methods, and Applications,"* Birkhauser, Boston, (2000).

Balakrishnan, N., Ng, H. K. T. and Kannan, N. "Goodness-of-fit tests based on spacings for progressively type-II censored data from a general location-scale distribution," *IEEE Trans. on Rel.,* **53**, 349–356, (200).

Balakrishnan, N., Habibi Rad, A. and Arghami, N. R. "Testing exponentiality based on Kullback-Leibler information with progressively type-II censored data," *IEEE Trans. on Rel.,* **53**, 349–356, (2007).

Dudewicz, E. J. and van der Meulen, E. C. "Entropy-based tests of uniformity," *J. Amer. Statist. Assoc.,* **76**, 967–974, (1981).

Ebrahimi, N. and Habibullah, M. "Testing exponentiality based on Kullback-Leibler information," *J. Royal Statist. Soc. B,* **54**, 739–748, (1992).

Gokhale, D. V. "On entropy-based goodness-of-fit tests," *Computational Statistics and Data Analysis,* **1**, 157–165, (1983).

Kullback, S. "*Information Theory and Statistics,"* New York: Wiley, (1959).

Park, S. "The entropy of consecutive order statistics," *IEEE Trans. Inform. Theory,* **41**, 2003–2007, (1995).

Park, S. "Testing exponentiality based on the Kullback-Leibler information with the type-II censored data," *IEEE Trans. on Rel.,* **54**, 22–26, (2005).

Shannon, C. E. "A mathematical theory of communications," *Bell System Tech. J.,* **27**, 379–423, (1948).

Vasicek, O. "A test for normality based on sample entropy," *J. Royal Statist. Soc. B,* **38**, 730–737, (1976).

Wong, K. M. and Chan, S. "The entropy of ordered sequences and order statistics," *IEEE Trans. Inform. Theory,* **36**, 276–284, (1990).

Weighted Averages With Random Proportions

H. Homei

Departments of Statistics, University of Tabriz

A weighted Average of n independent continuous random variables $X_1, \cdots, X_n$ with random proportions is introduced. A formula between the Stieltjes transforms of the weighted average and $X_1, \cdots X_n$ is expressed. In particular when $X_1, \cdots, X_n$ have a common distribution F, we show the weighted average is distributed as $X_1, \cdots, X_n$ if and only if F is a Cauchy distribution. Then we give examples for them.

1. Introduction

The work of Van Assche (1987) was commentated by Johnson and Kotz (1990). By considering the direct method of analysis based on the calculation of moments, they argued that simple and direct approaches sometimes appear to be more advantageous than advanced techniques, such as the Stieltjes transform technique employed by Van Assche (1987). They proceeded further and highlighted a derivation for the *Result (i),* from their method, but acknowledged that their approach was not useful to obtain the *Result (ii)*. Johnson and Kotz (1990) viewed the random variable S as a random weighted average of X and Y with random proportions U and $1-U$, $S = UX + (1-U)Y$, U uniform $[0,1]$ independent of X and Y. Soltani and Homei (2009) also expressed their interests in the extension of the Van Assche (1987) results when more than two random variables are involved, but In this work, we review the work of them.

1.1. *Definition*

Definition: Let $\mathbf{X} = (X_1, \cdots X_n)$ and $\mathbf{W} = (W_1, \cdots W_n)$, where $X_1, \cdots X_n$ be i.i.d with common distribution F and $W_1, W_2, ..., W_n$ are nonnegative random variables and subject to $\sum_{i=1}^{n} W_i = 1$, then

$$\bar{X}_W = W_1X_1 + W_2X_2 + \cdots + W_{n-1}X_{n-1} + W_nX_n,$$

$\bar{X}_W$ is defined to be RWA

Example 1: As especially case , if

$$P(W_1 = \frac{1}{n}, \cdots, W_n = \frac{1}{n}) = 1.$$

We get the sample mean, which is usually denoted by $\bar{x}$, that is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

the following example is a useful example in studding one dimensional disorder systems

Example 2: Let $\{(\alpha_1, \beta_1), (\alpha_2, \beta_2), ...\}$ be independent random vectors on the unite square $[0, 1] \times [0, 1]$ with common distribution function $F(x, y) = P(\alpha_i \leq x, \beta_i \leq y)$

With this sequence we can construct random 2×2 matrices

$$\begin{pmatrix} \alpha_n & 1-\alpha_n \\ \beta_n & 1-\beta_n \end{pmatrix}$$

and their products

$$\begin{pmatrix} \alpha_n & 1-\alpha_n \\ \beta_n & 1-\beta_n \end{pmatrix} = \begin{pmatrix} \alpha_n & 1-\alpha_n \\ \beta_n & 1-\beta_n \end{pmatrix} \begin{pmatrix} \alpha_{n-1} & 1-\alpha_{n-1} \\ \beta_{n-1} & 1-\beta_{n-1} \end{pmatrix} ... \begin{pmatrix} \alpha_1 & 1-\alpha_1 \\ \beta_1 & 1-\beta_1 \end{pmatrix}$$

Where α_i, β_i are independent. These matrices are stochastic since all the entries are positive and the elements of every row add up to unity. The product Occur in physics in study of one-dimensional disordered systems.

Notice that the sequence $\{(X_n, Y_n), n = 1, 2, ...\}$ can be obtained by a random difference equation

$$\begin{cases} X_n = \alpha_n X_{n-1} + (1-\alpha_n) Y_{n-1}, \\ \\ Y_n = \beta_n X_{n-1} + (1-\beta_n) Y_{n-1}, \ n = 2, 3... \end{cases}$$

It is very clear that $\{(X_n, Y_n), n = 1, 2, ...\}$ is a bivaiate Markov process.

Where X_n and Y_n are RWA.

Example 3: (General Regression Neural Networks) Let

$$Z_i = E[Z \mid x, y] + \varepsilon_i$$

where ε_i are independent, have zero mean, and have variance $\sigma^2(x, y)$. The conditional mean of Z given (x,y) known as a regression of Z on (x,y) is the solution minimizing mean squared error. If f(x,y,z) is the joint continuous probability density function then the conditional mean can be expressed by the following relation:

$$E[Z \mid x, y] = \frac{\int_{-\infty}^{\infty} z f(z \mid x, y) dz}{\int_{-\infty}^{\infty} z f(z \mid x, y)}$$

If the error is normally distributed and homoskedastic $\varepsilon_i \sim N(0, \sigma^2)$ The regression estimate becomes the best linear unbiased estimate in the maximum likelihood sense. Let us consider basic formulas concerning GRNN used in the study. This part of the work is based The density function f(x,y,z) can be estimated from the data by using nonparametric consistent estimators

$$f(x,y,z)=\frac{1}{2\pi^{1.5}h^3n}\sum_{i=1}^{n}exp(-2D_i^2/2h^2)exp[-(z-z_i)^2/2h^2]$$

where Gaussian kernel is used , and n is the number of measurements in training data set, h - is a bandwidth, and distance metric is:

$$D_i^2=(x-x_i)^2+(y-y_i)^2$$

Substituting the joint probability estimate into the conditional mean, gives the desired conditional mean Z given (x,y), also called Nadaraya-Watson kernel estimator (Hardle, 1989):

$$Z_m(X,Y)=\frac{\sum_{i=1}^{n}Z_iexp(-2D_i^2/2h^2)}{\sum_{i=1}^{n}exp(-2D_i^2/2h^2)}=\sum_{i=1}^{n}W_i(X,Y)Z_i$$

where the weights are:

$$\frac{exp(-2D_i^2/2h^2)}{\sum_{i=1}^{n}exp(-2D_i^2/2h^2)}=W_i(X,Y)$$

$$\sum_{i=1}^{n}W_i(X,Y)=1$$

That, $Z_m(X,Y)=\sum_{i=1}^{n}W_i(X,Y)Z_i$ is RWA.

2. RWA Formed On Observe Sample

This section has written for finding distribution RWA that an interesting kind of distribution has introduced. We follow conditional distribution for given distinct values $X_1=x_1,...,X_n=x_n$.

Theorem 1: The conditional distribution of the RWA $\bar{X}_U$, for given distinct values $X_1=x_1,...,X_n=x_n$ at z, $-\infty<z<+\infty$ where $x_1>x_2>...>x_n$ will be given by

$$\begin{cases} K(z|x_1,...,x_n)=\sum_{j=0}^{i}\frac{(z-x_{n-j})^{n-1}}{C(x_{n-i};\,x_1,...,x_n)}, \\ x_{n-i}<z\le x_{n-i-1},\ i=0,\cdots,n-2, \end{cases} \tag{1}$$

where

$$C(x_{n-j};\,x_1,...,x_n)=\prod_{k=1}^{n-j-1}(x_k-x_{n-j})\prod_{k=n-j+1}^{n}(x_k-x_{n-j}),\quad j=0,\cdots,n-1.$$

.

Example 4: As specially case $n=4$, conditional distribution function is

$$G(z)=\begin{cases} \frac{(z-x_4)^3}{x_{14}\,x_{24}\,x_{34}}, & x_4<z\le x_3 \\ \frac{(z-x_4)^3}{x_{14}\,x_{24}\,x_{34}}-\frac{(z-x_3)^3}{x_{13}\,x_{23}\,x_{34}}, & x_3<z\le x_2 \\ \frac{(z-x_4)^3}{x_{14}\,x_{24}x_{34}}-\frac{(z-x_3)^3}{x_{13}\,x_{23}x_{34}}+\frac{(z-x_2)^3}{x_{12}\,x_{23}\,x_{24}}, & x_2<z\le x_1. \end{cases}$$

where $x_{ij} = x_i - x_j$ and $x_4 < x_3 < x_2 < x_1$
Also when $n = 3$, we have

$$F(z) = \begin{cases} \frac{(z-x_3)^2}{x_{13}\ x_{23}}, & x_3 < z \leq x_2 \\ \\ \frac{(z-x_3)^2}{x_{13}\ x_{23}} - \frac{(z-x_2)^2}{x_{12}\ x_{23}}, & x_2 < z \leq x_1. \end{cases}$$

$$\frac{(z-x_3)^2}{x_{13}\ x_{23}} - \frac{(z-x_2)^2}{x_{12}\ x_{23}} = 1 - \frac{(x_1-z)^2}{x_{13}\ x_{12}} \quad x_2 < z \leq x_1.$$

Density function of distribution function $F(z)$ is:

$$f(z) = \begin{cases} \frac{2(z-x_3)}{x_{13}\ x_{23}}, & x_3 < z \leq x_2 \\ \\ \frac{2(x_1-z)}{x_{13}\ x_{12}}, & x_2 < z \leq x_1. \end{cases}$$

$f(z)$ is Two sided power distributions (soltani and homei (2009)).

Theorem 2: The conditional distribution of the $\bar{X}_U$, for given distinct values $X_1 = x_1, ..., X_n = x_n$ at z, $-\infty < z < +\infty$ will be given by (1).

3. distribution of RWA

In this section we present the main result of this chapter. we are going to profound theorem which according to that we are able to get Stieltjes transforms for RWA_U to Stieltjes transform for random variables $X_1, ..., X_n$. Therefore ,with distibutin of $X_1, ..., X_n$, we can find distribution of $[RWA]_U$ by the using of the Stieltjes transform theorem. This is following by unique property of Stieltjes transform.

Theorem 3: Under the assumption that $X_1, \cdots, X_n$ are independent and continuous,

$$\frac{(-1)^{n-1}}{(n-1)!}\frac{d^{n-1}}{dz^{n-1}}\mathcal{S}(F_{\bar{X}_U}, z) = \prod_{i=1}^{n}\mathcal{S}(F_{X_i}, z), \qquad z \in \mathbb{C}\bigcap_{i=1}^{n}(\text{supp}F_{X_i})^c. \tag{2}$$

Example 7: Let X_1, X_2 be i.i.d with Arcsin distribution on [-1,1], then

$$\mathcal{S}(X_i, z) = \frac{1}{\sqrt{1-z^2}}$$

By using theorem, we have

$$-\frac{d}{dz}\mathcal{S}(F_{\bar{X}_U}, z) = \frac{1}{z^2-1}$$

$$\begin{aligned}\mathcal{S}(F_{\bar{X}_U}, z) &= \int \frac{1}{1-z^2}dz \\ &= \frac{1}{2}\int \frac{1}{1-z}dz + \frac{1}{2}\int \frac{1}{1+z}dz \\ &= \frac{1}{2}(-\ln(z-1) + \ln(z+1)) \\ &= \int_{-1}^{1} \frac{1}{z-x}\frac{1}{2}dx \\ &= \mathcal{S}(F_{\bar{X}_U}, z)\end{aligned}$$

Where $F_{\bar{X}_U}$ is uniform distribution on [-1,1].

Example 8: Let X_1, X_2, X_3 be i.i.d with cauchy distribution , then

$$\mathcal{S}(X_i, z) = \frac{1}{z+c}$$

By using theorem, we have

$$\frac{(-1)^2}{2!}\frac{d^2}{dz^2}\mathcal{S}(F_{\bar{X}_U}, z) = (\frac{1}{z+c})^3$$

$$\begin{aligned}\mathcal{S}(F_{\bar{X}_U}, z) &= 2\int\int \frac{1}{(z+c)^3}dzdz \\ &= \frac{1}{z+c} \\ &= \int_{-\infty}^{\infty} \frac{1}{z-x}\frac{1}{\pi(1+x)}dx \\ &= \mathcal{S}(F_{\bar{X}_U}, z)\end{aligned}$$

Therefore, $F_{\bar{X}_U}$ is cauchy distribution.

Example 9: Let $X_1, \cdots, X_n$ be i.i.d with uniform distribution on [0,1], then

$$\begin{aligned}\mathcal{S}(X_i, z) &= \int_0^1 \frac{1}{(z-x)}dx \\ &= \ln(z-1) - \ln z\end{aligned}$$

By using theorem, we have

$$\frac{(-1)^{n-1}}{(n-1)!}\frac{d^{n-1}}{dz^{n-1}}\mathcal{S}(F_{\bar{X}_U}, z) = (\ln(z-1) - \ln z)^n$$

As specially case $n = 2$ we have

$$-\frac{d}{dz}\mathcal{S}(F_{\bar{X}_U}, z) = (\ln(z-1) - \ln z)^2$$

4. Conjugate

In the following of discussion about RWA , this question is defined that:

What distribution should we are use for $X_1, ..., X_n$ to has the same one for RWA ? Is cauchy that distribution we are looking for? To find the distribution $\bar{X}_W$ we can use the characterization function technique and double conditional expectation. Then Randomly Weighted Average has cauchy distribution for any n, that is, Randomly Weighted Average has the same distribution as one of its components.

Theorem 4: Let $\bar{X}_U$ be $[RWA]_U$ and $X_1, ..., X_n$ are independent and identically distributed continuous random variables with a common distribution function F. Then $\bar{X}_U$ has distribution F if and only if

$$\frac{(-1)^n}{(n-1)!}\frac{d^{n-1}}{dz^{n-1}}\mathcal{S}(F,z) = [\mathcal{S}(F,z)]^n, \quad z \in \mathbb{C}. \tag{3.6.1}$$

A solution for $\mathcal{S}(F,z)$ in (3.6.1) is

$$\mathcal{S}(F,z) = \frac{1}{z-a+ib}, \quad \mathrm{Im}(z) > 0, \quad b \neq 0,$$

which is the Stieltejes Transform of the Cauchy distribution.

5. Arc Sin Distribution

Theorem 5:$\bar{X}_U$ is uniform on $[-1,1]$ if and only if $X_1, ..., X_n$ have Stieltejes Transform

$$\mathcal{S}(F,z) = \left\{ \frac{1}{2(n-1)} \frac{(z+1)^{n-1} - (z-1)^{n-1}}{(z^2-1)^{n-1}} \right\}^{\frac{1}{n}}$$

References

1. Diniz, M.C., E. de Souza e Silva and H.R. Gail (2002). Calculating the distribution of a linear combination of uniform order statistics. Informs J. Computing, 14, 0124-0131.
2. Johnson N.L. and S. Kotz (1990). Randomly weighted averages: Some aspects and extensions. American Statistician, 44, 245-249.
3. Soltani A.R. And H. Homei (2009). A generalization for two-sided power distributions and adjusted method of moments. Statistics,43, 611 - 620.
4. Soltani A.R. And H. Homei (2009). Weighted averages with random proportions that are jointly uniformly distributed over the unit simplex . Statistics and Probability Letters, 9, 1215-1218.
5. Van Assche, W. (1987). A random variable uniformly distributed between two independent random variables. Sankhyā, Ser. A, 49, 207-211.
6. Weisberg, H. (1971). The distribution of linear combinations of order statistics from the uniform distribution. Ann. Math. Statist., 42, 704-709.
7. Zayed, A.I. (1996). Handbook of Function and Generalized Function Transformations. CRC Press, London.

Minimax Estimation in Bounded Parameter Space for some Discrete Distributions under LINEX Loss Function

N. Jafari Tabrizi and N. Nematollahi

Department of Mathematics, Karaj Branch, Islamic Azad University

Department of Statistics, Allameh Tabataba'i University

In many statistical problems, there exists bounds on the values that unknown parameters can take. In this paper in a class of discrete distributions including Bernoulli(θ), Binomial(k, θ), Discrete Wiebull(θ), ..., we consider minimax estimation of the parameter θ when it lies in a bounded interval of the form $[0, m]$ under LINEX loss function. We give conditions on m for which the Bayes estimator of $\theta \in [0, m]$ is minimax under LINEX loss function.

Keywords: Bayes estimator, Bounded parameter space, LINEX loss function, Minimax estimation.

1. Introduction

Estimation in bounded parameter space is one of the most important problems in statistical inference. While the assumption of boundedness of the parameter can be useful in practice, it introduce some surprising difficulties in theory. Initial attempts to investigate estimation of bounded parameter space were made by Hammersley (1950), in which the Maximum Likelihood (ML) estimators of Normal and Poisson means are considered. ML estimators are usually inadmissible in bounded parameter space. Minimax estimation is one of the common methods that used in bounded parameter space and usually leads to admissible estimators. For a review of admissible and minimax estimation in bounded parameter space see Marchand and Strawderman (2004) and van Eeden (2006).

Marchand and Parsian (2006) for a vast class of discrete distributions includes Binomial(k, θ) and Poisson(θ), give sufficient and necessary conditions for which the boundary supported Bayes estimator of $\theta \in [0, m]$ is minimax under Squared Error Loss (SEL) function. In this paper we consider a class of discrete distributions, including Binomial (k, θ), Discrete Wiebull (θ), Consul (k, θ),..., and derive minimax estimator of θ when it lies in a bounded interval of the form $\theta \in [0, m]$ under the LINEX loss function given by

$$L(\theta, \delta) = c\{e^{a(\delta-\theta)} - a(\delta - \theta) - 1\}, \quad c > 0, \quad a \neq 0. \tag{1}$$

The LINEX loss function was introduced by Varian (1975) and is convex in δ and is not

symmetric. This loss function is useful when overestimation is more (less) serious than underestimation of the same magnitude. For a review of estimation under LINEX loss function see Parsian and Kirmani (2002). clearly the value of $c > 0$ does not have any influence on our results, therefore without loss of generality, we shall take $c = 1$ in the rest of the paper.

In section 2 we state some preliminary results and introduce the class of discrete distributions. In section 3, we derive the least favorable prior and minimax estimator of $\theta \in [0, m]$ in the class of discrete distributions under LINEX loss function.

2. Preliminary Results and Class of Discrete Distributions

Let $\mathbf{X} = (X_1, \cdots, X_n)$, $n \geq 1$ be a set of discrete random variables with joint probability function $f(\mathbf{x}, \theta)$, where $\theta \in [0, m]$ for some $m > 0$. It follows from the work of DasGupta (1985) that the least favorable prior is quite generally supported on the boundary $\{0, m\}$ of the parameter space, and that the corresponding Bayes estimator is minimax for small enough m. So our results are based on the following well-known criteria for minimaxity applied to boundary two-point priors.

Lemma 2.1: A two-point boundary prior π on $\{0, m\}$ is least favorable, and the corresponding Bayes estimator $\delta_\pi(\mathbf{x})$ is minimax, if and only if

$$R(0, \delta_\pi) = R(m, \delta_\pi) = \sup_{0 \leq \theta \leq m} R(\theta, \delta_\pi). \tag{2}$$

For finding the equalizer rule, i.e., the Bayes rule which satisfy $R(0, \delta_\pi) = R(m, \delta_\pi)$, we use the following notations and result of Wan et al. (2000).

Let $\mathbf{X} = (X_1, ..., X_n)$ be a set of random variables and P_θ be the distribution of $\mathbf{X}$ with the parameter space $\theta \in [\nu, \beta]$, $\nu < \beta$. Assume that P_θ is dominated by some σ-finite measure μ. Further let $f(\mathbf{x}, \theta)$ be the Radon-Nikodym derivative of P_θ with respect to μ. Assume that:
(i) $f(\mathbf{x}, \nu) + f(\mathbf{x}, \beta) \neq 0$ for all $\mathbf{x} \in \mathtt{X}$, where $\mathtt{X}$ is the sample space, and
(ii) $P_\theta\{\mathbf{x} : f(\mathbf{x}, \nu) f(\mathbf{x}, \beta) > 0\} > 0$ when $\theta = \nu$ and $\theta = \beta$.

Consider the following two point prior

$$\pi(\nu) = \eta \quad , \quad \pi(\beta) = 1 - \eta, \tag{3}$$

where $0 < \eta < 1$. Following Wan et al. (2000), the corresponding Bayes estimator under the LINEX loss function (1) is

$$\delta_\pi(\mathbf{x}) = \frac{1}{a} \ln \left\{ \frac{\eta f(\mathbf{x}, \nu) + (1 - \eta) f(\mathbf{x}, \beta)}{\eta f(\mathbf{x}, \nu) e^{-a\nu} + (1 - \eta) f(\mathbf{x}, \beta) e^{-a\beta}} \right\}. \tag{4}$$

Note that $\delta_\pi(\mathbf{x}) = \beta$ if $f(\mathbf{x}, \nu) = 0$, and $\delta_\pi(\mathbf{x}) = \nu$ if $f(\mathbf{x}, \beta) = 0$.

Theorem 2.1 *(Wan et al., 2000). Under the LINEX loss function (1), there exists a unique* $\eta^* \in (0, 1)$ *such that* $R(0, \delta_{\pi^*}) = R(m, \delta_{\pi^*})$, *where* π^* *is the prior distribution*

for $\eta = \eta^*$. *Moreover,* $R(0, \delta_\pi) < R(m, \delta_\pi)$ *for* $\eta \in (\eta^*, 1)$ *and* $R(0, \delta_\pi) > R(m, \delta_\pi)$ *for* $\eta \in (0, \eta^*)$.

In the next section we use Theorem 2.1 to construct equalizer rule that satisfy minimaxity condition (2). Now, we introduce the class of discrete distributions that used in this paper.

Let $\mathbf{X} = (X_1, \cdots, X_n)$; $n \geq 1$ where $X_i's$ are identically but not necessarily independent distributed, discrete random variables with joint probability function $f(\mathbf{x}, \theta) = P_\theta(\mathbf{X} = \mathbf{x})$ and the support of $\mathbf{X}$ is lower bounded by $(\underline{s}, \ldots, \underline{s})$. We consider minimax estimation of θ under LINEX loss function when θ is bounded to a small enough known interval $[b, d] \subset \Theta$ where Θ is the unconstrained parameter space and the distribution of $\mathbf{X}$ under $\theta = b$ is degenerate at $(\underline{s}, \cdots, \underline{s})$. Since we can translate X_i to $X_i - \underline{s}$, we assume hereafter $\underline{s} = 0$. Also, since $\delta(\mathbf{X})$ is minimax for θ under the LINEX loss function (1) if and only if $\delta(\mathbf{X}) - b$ is minimax for $\theta - b$, so without loss of generality we assume hereafter that $[b, d] = [0, m]$.

Let $A = \{\mathbf{x} \in \mathbf{R}^n : \sum_{i=1}^{n} x_i = 0\}$ and $G(n, \theta) = P_\theta(\mathbf{X} \in A)$, then $G(n, 0) = 1$ by assumption. Marchand and Parsian (2006) and Jafari Tabrizi and Nematollahi (2009) considered the following classes of discrete distributions

$$C^* = \{f(., \theta) : G(n, 0) = 1, (-1)^k \frac{\partial^k}{\partial \theta^k} G(n, \theta) > 0 \ \text{ for } \ \theta \geq 0 \ \text{ and } \ k = 1, 2, 3\} \quad (5)$$

and

$$C_1 = \{f(., \theta) : G(n, \theta) = e^{\alpha(n)\theta}, \ \ \alpha(n) < 0, \ \ \theta \geq 0\} \subset C^*$$

respectively. They derive sufficient (and necessary) conditions for which the Bayes estimator of θ with respect to a boundary supported prior on $\{0, m\}$ be minimax under SEL and LINEX loss function, respectively. We are interested in minimax estimation of θ under the LINEX loss function (1) in a subclass C of C^* given by

$$C = \{f(., \theta) : G(n, \theta) = (1 - \theta)^{\alpha(n)}, \alpha(n) \geq 1, \theta \geq 0\} \quad (6)$$

where $\alpha(n)$ is a function of n and known parameters and does not depend on θ. Some discrete distributions that belong to the class C are:
1- $X_i's$ are independently distributed Bernoulli(θ), with $G(n, \theta) = (1 - \theta)^n$.
2- $X_i's$ are independently distributed Binomial(k, θ), with known k, $G(n, \theta) = (1 - \theta)^{kn}$.
3- $X_i's$ are independently distributed Negative Binomial($r, 1-\theta$), with known r, $G(n, \theta) = (1 - \theta)^{rn}$.
4- $X_i's$ are independently distributed Discrete Wiebull($1 - \theta$), with $G(n, \theta) = (1 - \theta)^n$.
5- $X_i's$ are independently distributed Generalized Negative Binomial(k, θ), with known k, $G(n, \theta) = (1 - \theta)^{kn}$.
6- $X_i's$ are independently distributed Consul(k, θ) with pf

$$f(x, \theta) = \frac{\Gamma(kx + 1)}{\Gamma(x + 1)\Gamma(kx - x + 2)} (\frac{\theta}{1 - \theta})^{x-1} (1 - \theta)^{kx}, \ x = 1, 2, \cdots, \ \ 0 < \theta \leq 1,$$

where $k \in \{1, 2, ...\}$ is known, $G(n, \theta) = (1-\theta)^{kn}$.

The above family of distributions and also some other distributions that belong to the class C can be found in Johnson et al. (2005).

3. Minimax Estimation of $\boldsymbol{\theta}$

Let $\mathbf{X} = (X_1, \cdots, X_n)$ has a joint probability function $f(., \theta)$ belong to class C of discrete distributions of the form (6). In this section we want to find minimax estimator of θ when $\theta \in [0, m]$. Since $P_\theta(\mathbf{X} \in A) = P_\theta(X_1 = 0, \cdots, X_n = 0) = G(n, \theta) > 0$ for all $\theta > 0$ and $G(n, 0) = 1$, therefore it can be easily seen that the conditions (i) and (ii) of Section 2 hold for $f(., \theta)$ in class C. So, we can use the results of Section 2 to obtain minimax estimator of θ.

Using two-point prior (3), the corresponding Bayes estimator (4) in class C is given by

$$\begin{aligned}\delta_\pi(\mathbf{x}) &= \begin{cases} \frac{1}{a}\ln\left\{\dfrac{\eta + (1-\eta)(1-m)^{\alpha(n)}}{\eta + (1-\eta)(1-m)^{\alpha(n)}e^{-am}}\right\} & \mathbf{x} = \mathbf{0} \\ m & \mathbf{x} \neq \mathbf{0} \end{cases} \\ &= yI_A(\mathbf{x}) + m(1 - I_A(\mathbf{x})) \end{aligned} \tag{7}$$

where $\mathbf{0} = (0, 0, \cdots, 0)$, $y = \frac{1}{a}\ln B$, $B = \dfrac{\eta + (1-\eta)(1-m)^{\alpha(n)}}{\eta + (1-\eta)(1-m)^{\alpha(n)}e^{-am}}$, and $I_A(.)$ is the indicator function. From Theorem 2.1, there exists a unique $\eta^* \in (0, 1)$ such that the Bayes estimator δ_{π^*} with respect to the two-point prior

$$\pi(0) = \eta^* \quad , \qquad \pi(m) = 1 - \eta^* \tag{8}$$

is equalizer Bayes rule, i.e., $R(0, \delta_{\pi^*}) = R(m, \delta_{\pi^*})$. To show that this equalizer Bayes rule is minimax, i.e., satisfy (2), we use convexity of $R(\theta, \delta_\pi)$ on $[0, m]$. Sufficient condition for convexity of $R(\theta, \delta_\pi)$ when $\alpha(n) \geq 1$ is given in the following theorem.

Theorem 3.1. *For $\alpha(n) \geq 1$ the risk function $\mathrm{R}(\theta, \delta_\pi)$ of the Bayes estimator δ_π is strictly convex on $[0, m]$ for every $\eta \in (0, 1)$, if one of the following conditions holds:*
(i) *$a \geq -\alpha(n)$ and $m \leq m_0$, where $m_0 \in (0, 1)$ is the unique root of the equation*

$$\begin{aligned}(-2a\alpha(n) - \alpha(n)(\alpha(n)-1))e^{am} + \alpha(n)(\alpha(n)-1)am \\ + (a + \alpha(n))^2 - \alpha(n) = 0 \end{aligned} \tag{9}$$

or
(ii) *$a < -\alpha(n)$ and $m \leq m_1$, where $m_1 \in (0, 1)$ is the unique root of the equation*

$$(a^2 + \alpha(n))e^{am} + \alpha(n)(\alpha(n)-1)am - \alpha(n) = 0 \tag{10}$$

Proof. From (7) the risk function of δ_π under the loss (1) is

$$\begin{aligned}\mathrm{R}(\theta, \delta_\pi) = [e^{a(m-\theta)} - a(m-\theta) - 1] + [e^{-a\theta}(B - e^{am}) \\ - (\ln B - am)](1-\theta)^{\alpha(n)}. \end{aligned} \tag{11}$$

The second derivative of $\mathrm{R}(\theta,\delta_\pi)$ with respect to θ is

$$\mathrm{R}''(\theta,\delta_\pi)=e^{-a\theta}(1-\theta)^{\alpha(n)-2}\Big\{\frac{a^2e^{am}}{(1-\theta)^{\alpha(n)-2}}+(B-e^{am})([a(1-\theta)+\alpha(n)]^2 - \alpha(n))-\alpha(n)(\alpha(n)-1)(\ln B-am)e^{a\theta}\Big\}. \quad (12)$$

(i) If $a>0$ then $1<B<e^{am}$ and $[a(1-\theta)+\alpha(n)]^2<(a+\alpha(n))^2$, therefore from (12) we have

$$\mathrm{R}''(\theta,\delta_\pi)\geq e^{-a\theta}(1-\theta)^{\alpha(n)-2}\{a^2e^{am}+(B-e^{am})[(a+\alpha(n))^2-\alpha(n)] - \alpha(n)(\alpha(n)-1)(\ln B-am)\}.$$

Let $\varphi_1(t)=[(a+\alpha(n))^2-\alpha(n)]t-\alpha(n)(\alpha(n)-1)\ln t$, then $\varphi_1(t)$ is a strictly convex function of t for $t>0$ and has a minimum at $t=\dfrac{\alpha(n)(\alpha(n)-1)}{(a+\alpha(n))^2-\alpha(n)}<1$. Hence $\varphi_1(B)>\varphi_1(1)=(a+\alpha(n))^2-\alpha(n)$, and

$$\begin{aligned}\mathrm{R}''(\theta,\delta_\pi)&>e^{-a\theta}(1-\theta)^{\alpha(n)-2}\big\{(-2a\alpha(n)-\alpha(n)(\alpha(n)-1))e^{am} + \alpha(n)(\alpha(n)-1)am+(a+\alpha(n))^2-\alpha(n)\big\}\\ &=e^{-a\theta}(1-\theta)^{\alpha(n)-2}\psi_1(m) \qquad (say). \end{aligned} \quad (13)$$

Note that $\psi_1'(m)=-2\alpha(n)a^2e^{am}+a\alpha(n)(\alpha(n)-1)(1-e^{am})<0$, hence $\psi_1(m)$ is strictly decreasing in m when $m>0$. Also $\lim\limits_{m\to0^+}\psi_1(m)=a^2>0$ and for $\alpha(n)\geq1$,

$$\begin{aligned}\lim_{m\to1^-}\psi_1(m)&=[-2a\alpha(n)-\alpha(n)(\alpha(n)-1)]e^a+\alpha(n)(\alpha(n)-1)a + (a+\alpha(n))^2-\alpha(n)\\ &\leq[-2\alpha(n)a-\alpha(n)(\alpha(n)-1)](a+1)+\alpha(n)(\alpha(n)-1)a + (a+\alpha(n))^2-\alpha(n)\\ &=a^2(1-2\alpha(n))<0. \end{aligned} \quad (14)$$

Therefore there exists a unique $m_0\in(0,1)$, the root of the equation (9), such that $\psi_1(m_0)=0$ and $\psi_1(m)>\psi_1(m_0)=0$ for $m<m_0$. Hence from (12), $R''(\theta,\delta_\pi)>0$ for $m\leq m_0$.

If $-\alpha(n)\leq a<0$ then $e^{am}<B<1$ and $[a(1-\theta)+\alpha(n)]^2>(a+\alpha(n))^2$. Therefore from (12) we have

$$R''(\theta,\delta_\pi)\geq e^{-a\theta}(1-\theta)^{\alpha(n)-2}\{a^2e^{am}+(B-e^{am})[(a+\alpha(n))^2-\alpha(n)] - \alpha(n)(\alpha(n)-1)(\ln B-am)\}.$$

Note that $\varphi_1'(t)=[(a+\alpha(n))^2-\alpha(n)]-\dfrac{\alpha(n)(\alpha(n)-1)}{t}$. If $0<(a+\alpha(n))^2<\alpha(n)<[\alpha(n)]^2$ then $\varphi_1(t)$ is strictly decreasing in t when $t>0$, and if $0<\alpha(n)<(a+\alpha(n))^2<[\alpha(n)]^2$ then $\varphi_1(t)$ is strictly decreasing in t when $0<t<1<\dfrac{\alpha(n)(n-1)}{(a+\alpha(n))^2-\alpha(n)}$. In

either case $\varphi_1(B) > \varphi_1(1) = (a+\alpha(n))^2 - \alpha(n)$. The remainder of the proof is similar to the proof of the case $a > 0$. See the appendix for a proof of $\lim_{m\to 1^-} \psi_1(m) \le 0$ when $-\alpha(n) \le a < 0$.

(ii) If $a < -\alpha(n)$ then $e^{am} < B < 1$, therefore from (12) we have

$$R''(\theta,\delta_\pi) \ge e^{-a\theta}(1-\theta)^{\alpha(n)-2}\{a^2e^{am} - \alpha(n)(B-e^{am}) - \alpha(n)(\alpha(n)-1)(\ln B - am)\}.$$

Let $\varphi_2(t) = -\alpha(n)t - \alpha(n)(\alpha(n)-1)\ln t$, then $\varphi_2(t)$ is strictly decreasing in t when $t > 0$. Hence $\varphi_2(B) > \varphi_2(1) = -\alpha(n)$, and

$$\begin{aligned} R''(\theta,\delta_\pi) &\ge e^{-a\theta}(1-\theta)^{\alpha(n)-2}\left\{(a^2+\alpha(n))e^{am} + \alpha(n)(\alpha(n)-1)am - \alpha(n)\right\} \\ &= e^{-a\theta}(1-\theta)^{\alpha(n)-2}\psi_2(m) \quad (say). \end{aligned} \tag{15}$$

Note that $\psi_2'(m) = a(a^2+\alpha(n))e^{am} + \alpha(n)(\alpha(n)-1)a < 0$, hence $\psi_2(m)$ is strictly decreasing in m when $m > 0$. Also $\lim_{m\to 0^+} \psi_2(m) = a^2$ and for $\alpha(n) \ge 1$, $\lim_{m\to 1^-} \psi_2(m) = (a^2+\alpha(n))e^a + \alpha(n)(\alpha(n)-1)a - \alpha(n) = g_1(a)$. Note that $g_1'(a) = (a+1)^2e^a + (\alpha(n)-1)e^a + \alpha(n)(\alpha(n)-1) > 0$, hence $g_1(a) < g_1(0) = 0$ for $a < 0$, which implies $\lim_{m\to 1^-} \psi_2(m) < 0$.

Therefore there exists a unique $m_1 \in (0,1)$, the root of the equation (10), such that $\psi_2(m_1) = 0$ and $\psi_2(m) > \psi_2(m_1) = 0$ for $m < m_1$. Hence from (15), $R''(\theta,\delta_\pi) > 0$ for $m \le m_1$.

Remark 3.1. It is easy to show that for $a < 0$ and $a \ne -\alpha(n)$, $\psi_1(m) > \psi_2(m)$ and on the boundary value $a = -\alpha(n)$, $\psi_1(m) = \psi_2(m)$. So for $a = -\alpha(n)$ we can compute m_0 and m_1 from either equations (9) or (10).

Remark 3.2. If $a > -\alpha(n)$ and $m > m_0$, then from the proof of Theorem 3.1.(i) we have $\psi_1(m) < \psi_1(m_0) < 0$ and

$$\lim_{\theta\to 0^+,\eta\to 1^-} R''(\theta,\delta_\pi) = (-2a\alpha(n) - \alpha(n)(\alpha(n)-1))e^{am} + \alpha(n)(\alpha(n)-1)am + (a+\alpha(n))^2 - \alpha(n) = \psi_1(m) < 0.$$

So there exists $\delta > 0$ and $\varepsilon > 0$ such that $R''(\theta,\delta_\pi) < 0$ for $\theta \in (0,\delta)$ and $\eta \in (1-\varepsilon, 1)$, i.e., $R(\theta,\delta_\pi)$ is strictly concave in $(0,\delta)$ and hence not convex on $[0,m]$ for $\eta \in (1-\varepsilon,1)$. Thus condition (i) of Theorem 3.1 is also a necessary condition for $R(\theta,\delta_\pi)$ to be strictly convex function of $\theta \in [0,m]$ for every $\eta \in (0,1)$. It seems that the condition (ii) of Theorem 3.1 is also a necessary for $R(\theta,\delta_\pi)$ to be strictly convex function of $\theta \in [0,m]$, but we can not prove it.

Now, since δ_π^* is equalizer Bayes rule with respect to the two point prior (8), and under the conditions (i) and (ii) of Theorem 3.1 $R(\theta,\delta_{\pi^*})$ is strictly convex on $\theta \in [0,m]$, therefore the condition (2) hold. So, we have the following result.

Theorem3.2 *Suppose either conditions (i) or (ii) of Theorem 3.1 holds, then the two point prior* $\pi^*(0) = \eta^*$, $\pi^*(m) = 1-\eta^*$ *is the least favorable prior, and the corresponding Bayes estimator*

$$\delta_{\pi^*}(\boldsymbol{x}) = \begin{cases} \dfrac{1}{a}\ln\left\{\dfrac{\eta^* + (1-\eta^*)(1-m)^{\alpha(n)}}{\eta^* + (1-\eta^*)(1-m)^{\alpha(n)}e^{-am}}\right\} & \boldsymbol{x} = \boldsymbol{0} \\ m & \boldsymbol{x} \neq \boldsymbol{0} \end{cases}$$

is the minimax estimator of θ.

Remark 3.3. The Bernoulli(θ) distribution is in class C with $\alpha(n) = n$. So, the minimax estimator of Jafari Tabrizi and Nematollahi (2009) is a special case of our minimax estimator.

4. Appendix

From(14)

$$\begin{aligned} \lim_{m\to 1^-}\psi_1(m) &= [-2a\alpha(n) - \alpha(n)(\alpha(n)-1)]e^a + \alpha(n)(\alpha(n)-1)a \\ &\quad + (a+\alpha(n))^2) - \alpha(n) \\ &= g(a) \qquad (say). \end{aligned}$$

For $-\alpha(n) \le a < 0$ and $\alpha(n) \ge 1$, $\frac{a+\alpha(n)}{\alpha(n)} \ge a+1$. Therefore

$$\begin{aligned} g'(a) &= [-2\alpha(n) - 2a\alpha(n) - \alpha(n)(\alpha(n)-1)]e^a + \alpha(n)(\alpha(n)-1) + 2(a+\alpha(n)) \\ &\ge -[2(a+\alpha(n)) + \alpha(n)(\alpha(n)-1)]e^a + \alpha(n)(\alpha(n)-1) + 2(a+\alpha(n)) \\ &= [2(a+\alpha(n)) + \alpha(n)(\alpha(n)-1)](1-e^a) \ge 0. \end{aligned}$$

Therefore $g(a)$ is increasing in a when $-\alpha(n) \le a < 0$, and hence $g(a) \le g(0) = 0$.

Acknowledgments

This work is a part of the project entitled by "Minimax estimator of upper bounded parameter of a family of discrete distributions under asymmetric loss function ". The authors are grateful to the research council of Islamic Azad University-Karaj Branch for financial support.

References

1. DasGupta, A. (1985). Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. *Sankhya* Ser. A **47**:326-332.
2. Hammersley, J. M. (1950). On estimation of restricted mean. *J. Royal Statist. Soc.* Ser. B **12**:192-229.

3. Jafari Tabrizi, N., Nematollahi, N., (2009). Minimax estimation of bounded parameter of some discrete distributions under LINEX loss function. Accepted to appear in *Commun. statist. Theory and Methods.*

4. Johnson, N.L., Kemp, A.W., Kotz, S., (2005). *Univariate discrete distributions.* John Willey and Sons, New Jersey.

5. Marchand, E., Parsian, A. (2006). Minimax estimation of a bounded parameter of a discrete distribution. *Statist. Probab. Lett.* **76**:547-554.

6. Marchand, E., Strawderman, W.E. (2004). Estimation in restricted parameter space: a review. Festschrift for Herman Rubin, *IMS Lecture Notes-Monograph* Series **45**:1-24.

7. Parsian, A. and Kirmani, S.N.U.A., (2002). Estimation under LINEX loss function. *In Handbook off applied Econometrics and statistical Inference, Eds.*

8. van Eeden, C. (2006). *Restricted parameter space estimation problems admissibility and minimaxity properties.* Springer, New York.

9. Varian, H. R. (1975). A Bayesian approach to real estate assessment. In: Fienberg, S. E., Zellner, A. (Eds.), *Studies in Bayesian Econometrics and Statistics in Honour of L. J. Savage.* North- Holland, Amsterdam, 195-208.

10. Wan, T., Zou, G., Lee, A. (2000). Minimax and Γ-minimax estimation for a poisson distribution under LINEX loss when the parameter space is restricted. *Statist. Probab. Lett.* **50**:23-32.

Maximize the Sharpe Ratio and Minimize a VaR

H. Jafarpour and R. Maller

Mathematical Department, Islamic Azad University-Shiraz branch

School of Finance and Application Statistics, Australian National University

In addition to its role as the optimal ex ante combination of risky assets for a risk-averse inves-tor, possessing the highest potential return-for-risk tradeoff, the tangency or Maximum Sharpe Ratio portfolio in the Markowitz [1952], [1991] procedure plays an important role in asset management, as it minimizes the probability that a future portfolio return falls below the risk-free or reference rate. This is a kind of Value at Risk (VaR) property of the portfolio. In this paper we demonstrate the way this VaR, and related quantities, vary along the efficient fron-tier, emphasizing the special role played by the tangency portfolio.

Keywords: Efficient portfolio, portfolio, Sharpe ratio, tangency portfolio, Value at Risk (VaR).

1. The "Maximum Sharpe Ratio" and "Tangency" Portfolios

Given a universe of $d \geq 2$ risky assets having raw return vector μ and excess mean return vector $\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}} - r\boldsymbol{i}$ (relative to a reference rate r), and returns covariance matrix $\boldsymbol{\Sigma}$, form a portfolio by taking an allocation, that is, a linear combination with coefficients given by a vector, $\boldsymbol{X_p}$, say, of the assets.[a] The "reference rate" could be the prevailing risk-free rate, if there is one, or some other benchmark rate, such as the expected market return for the period, etc. Let R_p be the *ex post* excess return achieved from this portfolio, after the portfolio has been in place for a speci-fied, fixed, time period. Suppose R_p has expectation μ_P and variance σ_p^2. Thus

$$E(R_p) = \mu_P = \boldsymbol{X_p'\mu} \quad \text{and} \quad Var(R_p) = \sigma_p^2 = \boldsymbol{X_p'\Sigma X_p} \tag{1}$$

Assume that, for all such portfolios, the standardized *ex post* excess returns

$$\frac{R_p - \mu_P}{\sigma_p}$$

areidentically distributed, having the same distribution as a random variable Z, say, where $E(Z) = 0$ and $Var(Z) = 1$.

[a]Here i denotes a d-vector, each of whose elements is one, and a prime will denote a vector or matrix transpose

Define the population Maximum Sharpe Ratio as

$$SR := \max_{i'x=1} \left(\frac{\boldsymbol{X}'\boldsymbol{\mu}}{\sqrt{\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}}} \right) \tag{2}$$

This quantity has long been used in portfolio theory and practice (Sharpe [1963]), either in an *ex ante* fashion, where it can be used to decide on an optimal allocation giving an optimal return-risk tradeoff, or *ex post*, as a portfolio performance evaluation tool. It plays a significant role in both discrete and continuous time finance, and is an object of interest in research right up to the present day (see, e.g., Christensen and Platen [2007]).

The maximization in (2) is over all portfolios satisfying the "total allocation" constraint $\boldsymbol{i}'\boldsymbol{X} = 1$, that a unit amount of resources is invested. There is no requirement that the components of the vector X be nonnegative, so short selling of assets is allowed. The ratio in (2) is maximized taking its sign into account, as advocated, e.g. by Sharpe [1994]; we are interested in maximiz-ing the actual (risk-adjusted) return - that is, a measure sensitive to losses, as well as to gains.[b]

In this paper, we consider an optimality property of the "Maximum Sharpe Ratio" portfolio, that is, the portfolio achieving the maximum value in Eq.(2), which it possesses with regard to "Value at Risk". The ideas are illustrated with a textbook example in the next section. The third section has some necessary background material. We then go on in the fourth section to develop some ideas regarding realized returns on efficient portfolios, which are illustrated with the same textbook data, and in the fifth section, we examine the performance of a spectrum of portfolios calculated from monthly data on US stocks prior to the October 1987 stock market crash, showing how the tangent portfolio, and various other selected portfolios, per-formed prior to, and on, the day of the crash.

To conclude this section we mention some further facts we will need, concerning the connection between the maximum Sharpe ratio and what we will call the "tangency" portfolio. The quantity SR in Eq.(2) is the maximum Sharpe ratio achievable from the d assets. In textbooks, and in applications, the corresponding portfolio is often found or illustrated by drawing a tangent line in the $(\sigma_p, \tilde{\mu}_p)$ plane from the point $(0, r)$ (where r is the risk-free or reference rate) to the efficient frontier constructed from $\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}$. The coordinates of this point, $(\sigma_p, \tilde{\mu}_p)$, say, give the location of the maximum Sharpe ratio portfolio in the $(\sigma_p, \tilde{\mu}_p)$ plane, and the slope of the tangent line gives the maximum Sharpe ratio available for any portfolio constructed from this universe of assets. The corresponding allocation vector $\boldsymbol{X_T}$ can be calculated from Equation (41) in Merton [1972].

[b] In some studies, the quantity in Eq. (2) is squared before the maximization is done. While this simplifies the algebra, it unrealistically ignores possible expected losses on the portfolio.

Merton [1972] showed further, however, that this procedure can be misleading or in error, since a tangency point producing a maximum Sharpe ratio need not in fact exist. He gave a necessary and sufficient condition for this to be the case (Theorem II in Merton [1972]). Of course a *maximum* value of the Sharpe ratio still exists (and is finite), but it has to be found by other means; see, e.g., the method outlined in Maller & Turkington [2002]. The probability calculation in (3) below uses only the existence of the maximum Sharpe ratio portfolio, how-ever calculated; it does not require the existence of a tangent point to the efficient frontier. Nevertheless, we shall continue to refer to the portfolio with maximum Sharpe ratio as the "tangent portfolio" whether or not such exists. For all the data considered in this paper, it turns out that the tangent portfolio does in fact exist, so no confusion should result from this.

2. Value at Risk Property of the Maximum Sharpe Ratio Portfolio

Let X_T be the allocation vector corresponding to the portfolio obtained as a result of the maximization in Eq. (2). As discussed in the previous section, we will refer to this as the "tangency portfolio". Recall that the excess mean return vector $\boldsymbol{\mu}$ equals $\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}} - r\boldsymbol{i}$, where $\tilde{\boldsymbol{\mu}}$ is the mean raw return vector, and r is the reference rate. We can write

$$SR = \frac{\tilde{\mu} - r}{\sigma_T} = \frac{\mu_T}{\sigma_T},$$

where $\mu_T = \boldsymbol{X_T'\mu}$, $\tilde{\mu} = \boldsymbol{X_T'\tilde{\mu}}$ and $\sigma_T = \boldsymbol{X_T'\Sigma X_T}$. The maximum Sharpe Ratio portfolio possesses a certain optimality property with respect to VaR, as the following simple calculation shows. For an arbitrary portfolio with allocation X_p, we have

$$SR \geq \left(\frac{\boldsymbol{X'\mu}}{\sqrt{\boldsymbol{X_p'\Sigma X_p}}} \right) = \frac{\mu_P}{\sigma_p}.$$

Letting R_T be the excess return on the tangency portfolio, we can calculate

$$\begin{aligned} P(R_p \leq 0) &= P\left(\frac{R_P - \mu_P}{\sigma_p} \leq -\frac{\mu_P}{\sigma_p} \right) \\ &= P\left(Z \leq -\frac{\mu_P}{\sigma_p} \right) \\ &\geq P(Z \leq -SR) \\ &= P\left(\frac{R_T - \mu_T}{\sigma_T} \leq -\frac{\mu_T}{\sigma_T} \right) \\ &= P(R_T \leq 0) \end{aligned} \tag{3}$$

It follows that

$$\min_{X_P} P(R_P \leq 0) \geq P(R_T \leq 0),$$

and the minimum is achieved for the tangency portfolio. Thus, an allocation of assets according to the tangency portfolio has the lowest probability of the investor receiving a return below the reference rate; in other words, it has the smallest VaR relative to this rate.

As portfolios move away from the maximum Sharpe Ratio allocation, this probability increases. We can illustrate the magnitude of this increase by plotting the probability for portfolios on the efficient frontier, that is, those having expected return and standard deviation $(\tilde{\mu}_P, \sigma_p)$, against σ_p, thus obtaining a representation of the way this VaR changes along the efficient frontier. We have to assume a distribution for Z, the standardized return, and for this we will consider a standard normal, as well as a t-distribution with 4 degrees of freedom. These represent extreme distributions between which returns distributions are likely to lie. While the normal distribution is often assumed for returns, especially over longer periods, it has been long recognized that returns distributions in reality are more heavy tailed and leptokurtic than the nor-mal distribution (Fama, [1965]; Embrechts et al. [1997]; Platen and Sidorowicz [2007]); therefore, we utilize a t-distribution with small degrees of freedom to simulate this feature of the data.

We illustrate the concepts using some data from Ruppert's text book [2004, p.150]. There are $d = 3$ assets for which the (raw) mean vector and covariance matrix are

$$\tilde{\boldsymbol{\mu}} = \begin{pmatrix} 0.08 \\ 0.03 \\ 0.05 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 0.30 & 0.02 & 0.01 \\ 0.02 & 0.15 & 0.03 \\ 0.01 & 0.03 & 0.18 \end{pmatrix}.$$

The efficient frontier for this example is shown on p.155 of that book. In Fig.1. we plot the function $P(R_P \leq 0) = P(Z \leq -\mu_{_P}/\sigma_p)$ for portfolios on the efficient frontier, as a function of the portfolio risk, σ_p. As expected, the curves have a minimum at the tangent point, and the curve for the t-distribution is higher than for the normal; the probability of a return below the risk-free rate is much higher for the heavier-tailed t-distribution.

A "Value at Risk" is usually thought of as a quantile below which a return falls with a specified (low) probability; thus, we should also consider $P(R_P \leq q)$, for values of q not equal to zero. It is not the case in general that this quantity is minimized for the Maximum Sharpe Ratio allocation, but by observation this seems to remain approximately true for q not too far from zero (recall that we are optimizing *excess* returns, relative to a benchmark). In Fig.2. , the probabilities of efficient portfolio returns lower than q are shown for various values of q. For example, from the lowest curve in Fig.2. (left plot) can be read that the probability of a future excess return less than -0.02 for the tangent point portfolio is approximately 0.441. Thus the tangency portfolio is expected to return more than 0.02 below the reference rate at most 44.1% of the time. For the t_4 distribution in the right plot,

on the other hand, such a loss happens ap-proximately with probability 0.445.

Although it is not necessarily the case that the minima of the curves in Fig.2. should occur at the tangent point (expected for the cases $q = 0$), in fact this happens for this data.

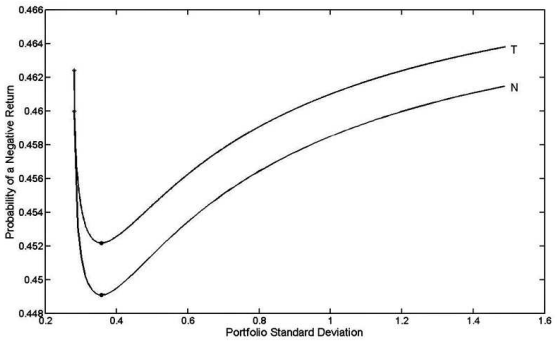

Fig. 1. Ruppert Data, Normally and t4 Distributed Returns

The probability of receiving a negative return, as a function of the standard deviation of the efficient portfolio. The curve labeled "N" depicts normally distributed returns and the curve labeled "T" depicts t4 distributed returns. The curves start at the standard deviation of the minimum variance portfolio on the left, and show the position of the tangent point portfolio (indicated by dot points).

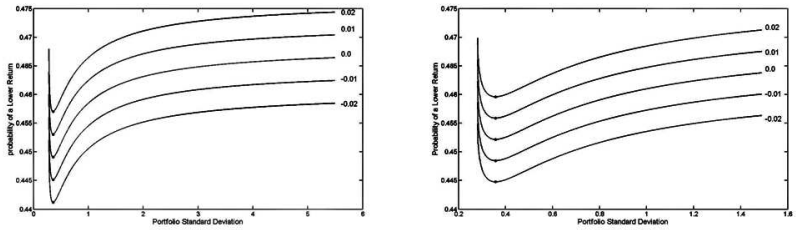

Fig. 2. Ruppert Data, Values at Risk for Normally and t4 Distributed Returns

The probability of receiving an excess return lower than q, where q is specified by the numbers at the right hand ends of the curves, as a function of the standard

deviation of the efficient portfolio. The left hand diagram depicts normally distributed returns; the right hand diagram t4 distributed returns. The curves start at the standard deviation of the minimum variance portfolio on the left, and show the position of the tangent point portfolio (indicated by dot points).

3. Efficient Portfolio Returns

To investigate the performances of portfolios on the efficient frontier, we need some facts concerning them. These are derived from Merton [1972]. In our notation, the quantities on p.1853 of his paper are:

$$A = \boldsymbol{i}^{'}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}, \quad B = \tilde{\boldsymbol{\mu}}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}, \quad C = \boldsymbol{i}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{i}, \quad D = BC - A^2 > 0$$

(Recall that $\tilde{\boldsymbol{\mu}}$ denotes the raw returns and $\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}} - r\boldsymbol{i}$ are the excess returns on the d assets.) We assume that a tangent point exists, so the quantity

$$TPC = \boldsymbol{i}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = A - rC$$

is positive (Merton[1972], p. 1863). The coordinates in the $(\tilde{\mu}, \sigma)$ plane of the minimum variance and tangent point portfolios are given by

$$\sigma_m^2 = \frac{1}{C}, \;\; \tilde{\mu}_m = \frac{A}{C}$$

and

$$\sigma_T^2 = \frac{\boldsymbol{\mu}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\left(\boldsymbol{i}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)^2}, \quad \tilde{\mu}_T = r + \frac{\boldsymbol{\mu}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{i}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}\,.$$

The corresponding portfolio allocations are

$$\boldsymbol{X}_m = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{i}}{C} \quad \text{and} \quad \boldsymbol{X}_T = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{i}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}\,.$$

(Note: we have $\boldsymbol{\mu}$, not $\tilde{\boldsymbol{\mu}}$, in σ_T, $\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{X}_T$.)
The equation of the efficient frontier in $(\sigma, \tilde{\mu})$ space is

$$F(\sigma) = \frac{A + \sqrt{D(C\sigma^2 - 1)}}{C} \tag{4}$$

(We use the "F" notation for "frontier", rather than Merton's "E" notation, which we reserve for "expectation".) The portfolio allocation corresponding to a portfolio with coordinates $(\sigma, \tilde{\mu})$ on the efficient frontier is given by the vector

$$X = \frac{F(\sigma)(C\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\mu}} - A\boldsymbol{\Sigma}^{-1}\boldsymbol{i}) + B\boldsymbol{\Sigma}^{-1}\boldsymbol{i} - A\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}}{D} \tag{5}$$

(Merton [1972], p.1856 and p.1845).

It is easily checked by differentiation that the curve

$$\frac{F(\sigma) - r}{\sigma} \tag{6}$$

in $(\sigma, \tilde{\mu})$ space has a maximum at the point σ_T which satisfies

$$\sigma_T^2 = \frac{B - 2rA + r^2C}{(A - rC)^2} \tag{7}$$

this of course is the variance of the tangent point portfolio. (Note that the denominator in Eq. (7) is $(TPC)^2 > 0$.) The function in Eq. (6) increases for $\sigma < \sigma_T$ and decreases for $\sigma > \sigma_T$. Fig. 3. shows the curve for the Ruppert data, taking $r = 0.02$ as on p.155 of Ruppert [2004].

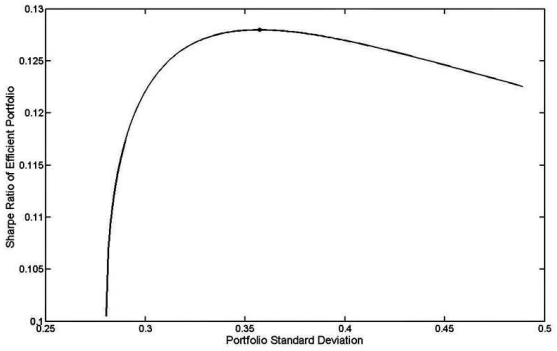

Fig. 3. Sharpe Ratios for Efficient Portfolios from the Ruppert Data

Plot of the Sharpe ratio (Eq. (2)) for portfolios on the efficient frontier, against their standard deviation, Ruppert textbook data. The tangent point portfolio is indicated by a dot.

Now suppose we have a new observation vector, $\tilde{R}$, on the returns of the d assets. We can think in terms of the efficient portfolio with mean $F(\sigma)$ and standard deviation σ being put in place at a certain time, then evaluated using the future return $\tilde{R}$. Using Eq. (5), and after some algebra, we can write

$$\frac{\tilde{\boldsymbol{R}}\boldsymbol{X} - r}{\sigma} = \frac{F(\sigma) - r}{\sigma} + \frac{(\tilde{\boldsymbol{R}} - \tilde{\boldsymbol{\mu}})((C - A)\boldsymbol{\Sigma^{-1}}\tilde{\boldsymbol{\mu}} + (B - A)\boldsymbol{\Sigma^{-1}}\boldsymbol{i})}{\sigma D}. \tag{8}$$

Here $\tilde{\boldsymbol{R}}X$ represents the return on the efficient portfolio corresponding to the returns $\tilde{\boldsymbol{R}}$ on the d assets, and the quantity on the left of Eq. (8) is the standardized excess return, i.e., the *ex post* Sharpe ratio for the portfolio. On the right of Eq. (8) is the population Sharpe ratio for the port-folio plus a random term corresponding to the new return, $\tilde{\boldsymbol{R}}$. If $\tilde{\boldsymbol{R}}$ is drawn from the same population as that from which

the efficient portfolio was constructed, so that $E(\tilde{\boldsymbol{R}}) = \tilde{\boldsymbol{\mu}}$ and $Var(\tilde{\boldsymbol{R}}) = \boldsymbol{\Sigma}$, it is clear that the expectation of the random term in Eq. (8) is zero, and its variance is one (as can also be checked after some algebra). Fig. 4. shows a plot of Eq. (8) for 13 returns generated randomly as observations on $N(\tilde{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$, using Ruppert's values of $\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}$. (Ruppert does not supply the original returns for which his $\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}$ were calculated, so we simulated the observations.) We took 13 returns so as to correspond with the 1987 crash data in the next section. It is clear from Fig. 4. that Eq. (8), as a function of σ, need not resemble Eq. (6), as shown plotted in Fig. 3. For this data, the random component in Eq. (8), which has a standard deviation of one, overwhelms its expectation, which for this data peaks at about 0.13 (cf. Fig. 3.).

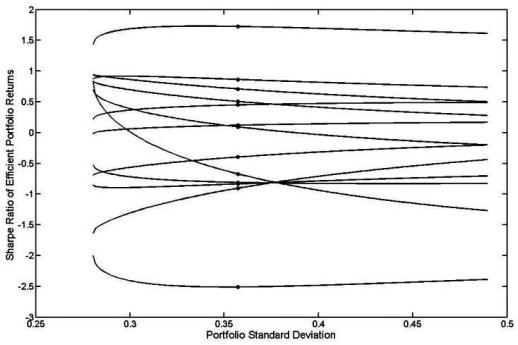

Fig. 4. Standardized Returns on Efficient Portfolios for Ruppert Data

Ex post Sharpe ratios for returns on portfolios on the efficient frontier, corresponding to a new return, against their standard deviations, Ruppert textbook data. The tangent point portfolio is indicated by a dot.

While the simulated future return curves sometimes peak close to the tangency point, at other times the maximum occurs for much higher risk portfolios, and sometimes the curves are even convex. For such data (and the data in the next section has similar features), unfortunately, investing in the tangency portfolio produces very little benefit for *individual* future returns. Only when averaged over a relatively large number of returns will curves calculated from Eq.(8) begin to resemble those from Eq.(6).

Discussion and Conclusion

Maximizing the return to risk trade-off through investing in the tangency portfolio is very well-known and understood by educated investors. Rational investors, especially investors whose trustees focus only on returns, will want to guard at all cost against the possibility that their portfolio will earn less than the risk-free, or reference, rate. Our work demonstrates the way that maximizing the expected Sharpe ratio through selecting the tangency portfolio minimizes the chances, not only of a return lower than the reference rate, but of even lower returns as well, across the range of efficient portfolios. These VaR minimizing properties of the tangency portfolio have not, to our knowledge, been implemented in a practical situation, and, as a result, the very desirable consequences of implementing a simple "black-box" approach to portfolio selection have not been thoroughly explored.

References

1. Christensen, Morton M., and Eckhard Platen. "Sharpe Ratio Maximization and Expected Utili-ty when Asset Prices Have Jumps", International Journal of Theoretical and Applied Finance, 10 (2007), pp. 1339 - 1364.
2. Embrechts, Paul, Claudia Klppelberg, Thomas Mikosch. Modelling Extremal Events For In-surance and Finance, Berlin: Springer-Verlag, 1997.
3. Fama, Eugene, F. "The Behavior of Stock-Market Prices." Journal of Business 38 (1965), pp. 34-105.
4. Maller, Ross A., Darrell A. Turkington. "New light on the portfolio allocation problem." Ma-thematical Methods of Operational Research, 56 (2002), pp. 501-511.
5. Markowitz, Harry. "Portfolio Selection." Journal of Finance, 7 (1952), pp. 77-91.
6. Markowitz, Harry. Portfolio Selection: Efficient Diversification of Investment, Cambridge, Mass.: Blackwell, 1991.
7. Merton, Robert C. "An analytic derivation of the efficient portfolio frontier." Journal of Finan-cial and Quantitative Analysis , 7 (1972), pp. 1851-1872.
8. Platen, Eckhard, R. Sidorowicz. "Empirical evidence on Student-t log-returns of diversified world indices." Journal of Statistical Theory and Practice, 2 (2007), pp. 1559-8608.
9. Ruppert, D. Statistics and Finance. An Introduction. Springer: New York, 2006.

Point Prediction of Order Statistics from Uniform Distribution

S. M. T. Kamel Mirmostafaee and J. Ahmadi

Department of Statistics, Ferdowsi University of Mashhad

In this paper, we consider the prediction problem in two sample while the sampling distribution is one parameter uniform distribution. Several approaches are discussed in order to find point predictors for future order statistics in a future sample on the basis of record statistics coming from the same distribution. At the end, a simulation study is given for illustrating the proposed procedures.

Keywords: Beta distribution, Order statistics, Point prediction, Record values.

1. Introduction

Let $X_1, \cdots, X_m$ be a random sample, if these random variables are arranged in ascending order of magnitude and then written as $X_{1:m} \leq X_{2:m} \leq \cdots \leq X_{m:m}$, then $X_{i:m}$ is said to be the ith order statistic in a sample of size m. Order statistics and functions of these statistics play an important role in a wide range of theoretical and practical problems such as characterization of probability distributions and goodness-of-fit tests, entropy estimation, analysis of censored samples, reliability analysis, quality control and strength of materials; see David and Nagaraja (2003) Arnold *et al.* (2008) and the references therein for more details.

Next, let $X_1, X_2, \cdots$ be an infinite sequence of random variables. Then, an observation X_j is said an upper record value if it exceeds all its previous observations, i.e., if $X_j > X_i$ for every $i < j$. Record data arise in a wide variety of practical situations including industrial stress testing, meteorological analysis, hydrology, seismology, sporting and athletic events, and oil and mining surveys. Interested readers may refer to the book by Arnold et al. (1998) and the references therein.

Several authors have studied the problem of predicting future order statistics based on observed order statistics in parametric or nonparametric models and also prediction future records on the basis of observed records from the same distribution. See for example, Lawless (1977), Dunsmore (1983), Raqab and Nagaraja (1995), Ahmadi et al.(2005), Ahmadi and Doostparast (2006) and the references therein. Recently, Ahmadi and Balakrishnan (2009) have obtained several nonparametric prediction intervals for future order statistics (record values) from

an independent Y-sequence based on order statistics (records) from X-sequence. Also, Ahmadi and Mirmostafaee (2009) have discussed the problem of predicting future order statistics based on observed usual records from an exponential distribution.

Let us denote the first n upper record values by $\mathbf{R} = (R_1, \ldots, R_n)$, and suppose the parent distribution is uniform $(0, \sigma)$ with probability density function(pdf)

$$f(x;\sigma) = \frac{1}{\sigma}, \quad 0 \leq x \leq \sigma. \tag{1}$$

Note that (1) belongs to the scale family of distributions with pdf $\frac{1}{\sigma} f(\frac{x}{\sigma})$. The main aim of this paper is predicting the future order statistics from (1) in terms of $\mathbf{R}$. In this regard, first we obtain best linear unbiased predictors(BLUPs) of the jth order statistic of a future sample of size m based on $\mathbf{R}$ in section 2. In section 3, we derive best linear invariant predictors(BLIPs). In section 4, another method is proposed for predicting order statistics by substituting the maximum likelihood estimator(MLE) of σ which is obtained from observed record data, to the expectation. Finally, in section 5, we use a simulated data set to present some numerical results for illustrating all the inferential methods developed here.

2. Best Linear Unbiased Predictor

In this section, we obtain the BLUP of the jth order statistic of a future sample of size m, i. e. $Y_{j:m}$. From the results on the generalized linear model due to Goldberger (1962), one can obtain the BLUP of $Y_{j:m}$ from a scale-parameter distribution as follows

$$Y^*_{j:m} = \sigma^* \alpha_{j:m} + \underline{\omega}^T \mathbf{\Sigma}^{-1} (\mathbf{R}^T - \sigma^* \underline{\alpha}), \tag{2}$$

where σ^* is the best linear unbiased estimator(BLUE) of σ, $\alpha_{j:m}$ equals $\frac{1}{\sigma} E(Y_{j:m})$, $\mathbf{R} = (R_1, \ldots, R_n)$, $\underline{\alpha}$ is the vector of means of record values from the standard distribution, $\mathbf{\Sigma}$ is the variance-covariance matrix of standard record values and

$$\underline{\omega}^T = \frac{1}{\sigma^2} (\mathrm{cov}(R_1, Y_{j:m}), \ldots, \mathrm{cov}(R_n, Y_{j:m})).$$

In our plan, we have two sample prediction, so $\mathbf{R}$ and $Y_{j:m}$ are independent which yields $\omega^T = (0, 0, \ldots, 0)$ and (2) simplifies as

$$Y^*_{j:m} = \sigma^* \alpha_{j:m}.$$

Here, as mentioned before, we consider (1) as the underlying distribution. It is well-known that the jth order statistic of a sample of size m from a standard uniform distribution, denoted by $U_{j:m}$, has beta distribution with parameters j

and $(m-j+1)$, i.e. $U_{j:m} \sim Beta(j, m-j+1)$. Noting that $\frac{1}{\sigma}Y_{j:m} \stackrel{d}{=} U_{j:m}$, we conclude $\alpha_{j:m} = \frac{j}{m+1}$. Also the BLUE of σ on the basis of $\mathbf{R}$ is

$$\sigma^* = \frac{4}{3^n - 1}\left[R_1 + \sum_{i=2}^{n} 3^{i-1}(R_i - \frac{1}{2}R_{i-1})\right],$$

and its variance is $2\sigma^2(3^{n+1}-3)^{-1}$ [see for example Arnold et. al. (1998)]. So by substituting, the BLUP of $Y_{j:m}$ based on observed $\mathbf{R}$ is given by

$$Y^*_{j:m} = \frac{4j}{(3^n-1)(m+1)} \sum_{i=1}^{n} 3^{i-1} W_i, \tag{3}$$

where

$$W_i = \begin{cases} R_1 & \text{for } \ i = 1, \\ R_i - \frac{1}{2}R_{i-1} & \text{for } \ i = 2, 3, \ldots, n. \end{cases} \tag{4}$$

Using the fact that $Y^*_{j:m}$ and $Y_{j:m}$ are independent and $\frac{1}{\sigma}Y_{j:m} \sim Beta(j, m-j+1)$, the mean squared prediction error(MSPE) of $Y^*_{j:m}$ is given by

$$\begin{aligned} \text{MSPE}(Y^*_{j:m}) &= E\left(Y^*_{j:m} - Y_{j:m}\right)^2 \\ &= \text{Var}(Y^*_{j:m}) + \text{Var}(Y_{j:m}) \\ &= \frac{j\sigma^2}{(m+1)^2}\left(\frac{2j}{3(3^n-1)} + \frac{m-j+1}{m+2}\right), \end{aligned}$$

which is decreasing with respect to n and m when other components are kept fixed.

3. Best Linear Invariant Predictor

We discussed the best linear unbiased prediction in the previous section. In this section, we focus our attention on linear predictors that are invariant with respect to location and scale transformations and also have the minimum MSPE among all invariant linear predictors. From the results of Mann (1969) and Kaminsky et al. (1975), we can get the BLIP of $Y_{j:m}$ from a scale-parameter distribution based on $\mathbf{R}$ as

$$\widetilde{Y}_{j:m} = Y^*_{j:m} - \left(\frac{V}{1+V}\right)\left(\alpha_{j:m} - \underline{\omega}^T \mathbf{\Sigma}^{-1} \underline{\alpha}\right)\sigma^*, \tag{5}$$

where $\underline{\omega}$, $\mathbf{\Sigma}$ and $\underline{\alpha}$ are as given in section 2 and $\sigma^2 V = \text{Var}(\sigma^*)$. By independency of $Y_{j:m}$ and $\mathbf{R}$, we have $\underline{\omega}^T = (0, 0, \ldots, 0)$. If we consider (1) as the underlying

distribution, by using the results of the previous section and simplifying them, we obtain

$$\widetilde{Y}_{j:m} = \frac{12j}{(3^{n+1}-1)(m+1)} \sum_{i=1}^{n} 3^{i-1} W_i, \tag{6}$$

where W_is are defined in (4). $\widetilde{Y}_{j:m}$ is not an unbiased predictor as we have

$$E(\widetilde{Y}_{j:m}) = \frac{3j(3^n-1)\sigma}{(3^{n+1}-1)(m+1)} \neq E(Y_{j:m}).$$

By (5), we have $\widetilde{Y}_{j:m} = Y^*_{j:m} - 2\alpha_{j:m}(3^{n+1}-1)^{-1}\sigma^*$, so using this relation we obtain

$$\begin{aligned} \mathrm{MSPE}(\widetilde{Y}_{j:m}) &= E\left(Y^*_{j:m} - 2\alpha_{j:m}(3^{n+1}-1)^{-1}\sigma^* - Y_{j:m}\right)^2 \\ &= \mathrm{MSPE}(Y^*_{j:m}) - \frac{4j^2\sigma^2}{3(m+1)^2(3^{n+1}-1)(3^n-1)} \\ &= \frac{j\sigma^2}{(m+1)^2}\left(\frac{2j}{3^{n+1}-1} + \frac{m-j+1}{m+2}\right), \end{aligned} \tag{7}$$

which is decreasing with respect to n and m when other components are considered to be fixed. From (7), we see that $\mathrm{MSPE}(\widetilde{Y}_{j:m}) < \mathrm{MSPE}(Y^*_{j:m})$. From the results due to Mann (1969), the best linear invariant estimator(BLIE) of σ is

$$\widetilde{\sigma} = \frac{\sigma^*}{1+V} = \frac{12}{(3^{n+1}-1)} \sum_{i=1}^{n} 3^{i-1} W_i,$$

where V is as before and W_is are defined in (4). By plugging this estimator into $E(Y_{j:m}) = \frac{j}{m+1}\sigma$, we arrive at same results for $\widetilde{Y}_{j:m}$ as in (6).

4. Plug-in the MLE of σ into the Expectation of $Y_{j:m}$

In this section, we propose other kinds of predictors for the jth order statistic in a future sample of size m. As we mentioned before $\frac{1}{\sigma}Y_{j:m} \sim Beta(j, m-j+1)$, so

$$E(Y_{j:m}) = \frac{j}{m+1}\sigma. \tag{8}$$

We suggest plugging the MLE of σ based on observed records into (8) in order to obtain a predictor for $Y_{j:m}$. Note that if we plug the BLUE and BLIE of σ into (8), the BLUP and BLIP of $Y_{j:m}$ will be obtained , respectively as in (3) and (6). Thus, it is reasonable to use other kinds of estimators of σ like its MLE to find other kinds of predictors for $Y_{j:m}$. The MLE of σ based on $\mathbf{R}$ is $\widehat{\sigma} = R_n$ [see e.g. Arnold et. al. (1998)], so a predictor for $Y_{j:m}$ can be given by

$$\widehat{Y}_{j:m} = \frac{j}{m+1} R_n,$$

which is simple in comparison with the BLUP and BLIP of $Y_{j:m}$. Clearly this predictor is not an unbiased one as $E(\hat{\sigma}) = (1-2^{-(n+1)})\sigma$. An unbiased estimator which can be used instead of $\hat{\sigma}$ is $\hat{\sigma}_u = (1-2^{-(n+1)})^{-1}R_n$. So an unbiased predictor for $Y_{j:m}$ becomes

$$\widehat{Y}^u_{j:m} = \frac{j}{(m+1)\left(1-2^{-(n+1)}\right)}R_n,$$

Using the knowledge that $\mathrm{Var}(\hat{\sigma}) = (3^{-(n+1)} - 4^{-(n+1)})\sigma^2$, [see Arnold et. al. (1998)], the MSPE of $\widehat{Y}_{j:m}$ is computed to be

$$\begin{aligned}
\mathrm{MSPE}(\widehat{Y}_{j:m}) &= E(\widehat{Y}_{j:m} - Y_{j:m})^2 \\
&= \left(\frac{j}{m+1}\right)^2 E(R_n^2) + E(Y^2_{j:m}) - 2\frac{j}{m+1}E(Y_{j:m}R_n) \\
&= \frac{j\sigma^2}{(m+1)^2}\left(\frac{j}{3^{n+1}} + \frac{m-j+1}{m+2}\right) \\
&= \mathrm{MSPE}(\widetilde{Y}_{j:m}) - \frac{j^2\sigma^2(3^{n+1}+1)}{(m+1)^2 3^{n+1}(3^{n+1}-1)}. \qquad (9)
\end{aligned}$$

From (9) we conclude that $\mathrm{MSPE}(\widehat{Y}_{j:m}) < \mathrm{MSPE}(\widetilde{Y}_{j:m})$, note that $\widehat{Y}_{j:m}$ is not a location invariant predictor, so this result is not unexpected. The MSPE of $\widehat{Y}^u_{j:m}$ is given by

$$\begin{aligned}
\mathrm{MSPE}(\widehat{Y}^u_{j:m}) &= E(\widehat{Y}^u_{j:m} - Y_{j:m})^2 \\
&= \mathrm{Var}(\widehat{Y}^u_{j:m}) + \mathrm{Var}(Y_{j:m}) \\
&= \frac{j\sigma^2}{(m+1)^2}\left(\frac{[3^{-(n+1)} - 4^{-(n+1)}]j}{(1-2^{-(n+1)})^2} + \frac{m-j+1}{m+2}\right).
\end{aligned}$$

It can be shown that for $n \geq 1$

$$3^{n+1} > 2^{n+2} - 1,$$

which is equivalent with

$$\frac{[3^{-(n+1)} - 4^{-(n+1)}]}{(1-2^{-(n+1)})^2} < \frac{1}{3^{n+1}}.$$

Therefore $\mathrm{MSPE}(\widehat{Y}^u_{j:m}) < \mathrm{MSPE}(\widehat{Y}_{j:m})$. Finally, from the above result, (7) and (9) we have

$$\mathrm{MSPE}(\widehat{Y}^u_{j:m}) < \mathrm{MSPE}(\widehat{Y}_{j:m}) < \mathrm{MSPE}(\widetilde{Y}_{j:m}) < \mathrm{MSPE}(Y^*_{j:m}). \qquad (10)$$

5. Illustrative Example

In this section, we have utilized a simulated set of record data in order to illustrate the procedures developed in this paper. The first $n = 5$ upper records were simulated from a one parameter uniform with pdf (1) with scale parameter $\sigma = 10$ using MINITAB version 11. They are as follows:

$$2.28,\ 4.75,\ 8.65,\ 9.12,\ 9.22.$$

Suppose we are interested in predicting some of the order statistics of a future sample of size 5, 10 and 20 in terms of these observed record values from the parent distribution, i.e. $U(0, 10)$. The BLUE, BLIE and MLE of σ are 9.53, 9.50 and 9.22, respectively. Table 1 contains the point predictors of some selected order statistics of a future sample of size m, $(m = 5, 10, 20)$ and their MSPEs based on the procedures discussed in this paper.

Table 1. Point predictors of some selected order statistics of a future sample of size m, $(m = 5, 10, 20)$.

m	j	$Y^*_{j:m}$ (BLUP)	$\widetilde{Y}_{j:m}$ (BLIP)	$\widehat{Y}_{j:m}$	$Y^u_{j:m}$	MSPE($Y^*_{j:m}$)	MSPE($\widetilde{Y}_{j:m}$)	MSPE($\widehat{Y}_{j:m}$)	MSPE($Y^u_{j:m}$)
5	1	1.59	1.58	1.54	1.56	1.9918	1.9918	1.9917	1.9874
	4	6.35	6.34	6.15	6.24	3.2970	3.2967	3.2965	3.2263
	5	7.94	7.92	7.68	7.81	2.1754	2.1749	2.1746	2.0649
10	1	0.87	0.86	0.84	0.85	0.6910	0.6910	0.6910	0.6897
	4	3.47	3.46	3.35	3.41	1.9648	1.9647	1.9647	1.9438
	8	6.93	6.91	6.71	6.81	1.7986	1.7982	1.7980	1.7144
	10	8.67	8.64	8.38	8.51	0.9164	0.9158	0.9154	0.7849
20	1	0.45	0.45	0.44	0.45	0.2068	0.2068	0.2068	0.2064
	4	1.82	1.81	1.76	1.78	0.7109	0.7109	0.7108	0.7051
	10	4.54	4.53	4.39	4.46	1.1963	1.1961	1.1960	1.1602
	17	7.71	7.69	7.46	7.58	0.8814	0.8809	0.8807	0.7771
	20	9.08	9.05	8.78	8.92	0.4560	0.4553	0.4550	0.3117

Although (10) is confirmed by Table 1(note that rounding error made the MSPEs seem equal for a few cases), the differences between MSPE($Y^*_{j:m}$), MSPE($\widetilde{Y}_{j:m}$), and MSPE($\widehat{Y}_{j:m}$) are not noticeable. However MSPE($\widehat{Y}^u_{j:m}$) is sensibly less than the others. It is also observed that all the MSPEs are decreasing with respect to m when other components are kept fixed but do not have a stable behavior with respect to j.

References

1. Ahmadi, J. and Balakrishnan, N. (2009), Prediction of order statistics and record values from two independent sequences, to appear in *Statistics*, DOI: 10.1080/02331880903189133.
2. Ahmadi, J. and Doostparast, M. (2006), Estimation and prediction in a two-parameter exponential distribution based on k-record values under LINEX loss function, *Statist. Papers*, 47, no.3, 373-392.
3. Ahmadi, J.; Doostparast, M. and Parsian, A. (2005), Bayesian estimation and prediction for some life distributions based on record values,*Comm. Statist. Theory Methods*, 34, no.4, 795-805.

4. Ahmadi, J. and MirMostafaee, S. M. T. K. (2009), Prediction intervals for future records and order statistics coming from two parameter exponential distribution, *Statistics & Probability Letters*, 79, 977-983.
5. Arnold, B. C.; Balakrishnan, N. and Nagaraja, H. N. (1998), *Records*, John Wiley & Sons, New York.
6. Arnold, B. C.; Balakrishnan, N. and Nagaraja, H. N. (2008), *A First Course in Order Statistics*, Classics in Applied Mathemathics, Number 54, society for Industrial and Applied Mathematics, Phliadelhia.
7. David H. A. and Nagaraja, H. N. (1998), *Order Statistics*, 3rd edition, John Wiley & Sons, New York.
8. Dunsmore, I. R. (1983), The future occurrence of records, *Ann. Inst. Statist. Math.*, 35, 276-277.
9. Goldberger, A. S. (1962), Best linear unbiased prediction in the generalized regression model,*J. Amer. Statist. Ass.* , 57, 369-375.
10. Kaminsky, K. S.; Mann N. R., and Nelson, P. I. (1975), Best and simplified linear invariant prediction of order statistics in location and scale families, *Biometrika*, 62, 525-526.
11. Lawless, J. F. (1977), Prediction intervals for the two parameter exponential distribution, *Technometrics* , 19, 469-472.
12. Mann, N. R. (1969), Optimum estimators for linear functions of location and scale parameters, *Annals of Mathematical Statistics*, 40, 2149-2155.
13. Raqab, M. Z. and Nagaraja, H. N. (1995), On some predictors of future order statistic, *Metron* , 53, 185-204.

Common Models in Accelerated Life Testing

M. Karbassian, M.H. Ghezel Ayagh and M. Bassel Helweh
Department of Industrial Engineering, University of Malek_Ashtar

Accelerated life tests (ALTs) are frequently used in practice to obtain failure time data quickly under high stress levels in order to predict product life performance under design stress conditions. Most of the previous works on ALT were concerned with the case where a single stress is employed for acceleration (For an excellent review, this is referred to Nelson [1] or Meeker and Escobar [2]).However, as components or products become more reliable due to advanced technology, it has become more difficult, even with a single-stress ALT, to obtain a sufficient amount of failure-time data within a reasonable amount of time.Multiple stress have been employed as a means of overcoming such difficulties.
In this paper; we discuss different models for accelerated life tests with two stresses that relate the failure data at stress conditions with design or operating conditions.

Keywords: Accelerated Failure Time Model, Proportional Hazards Model, Extended Hazard Regression Model,Proportional Mean Residual Life Model, Proportional Odds Models

1. introduction

Accelerated life testing (ALT) is a widely used approach for reliability demonstration and prediction of components or system reliabilities at normal operating conditions using data obtained at accelerated condition. Conducting an ALT requires the determination or development of a reliability inference model that relates the failure data at stress conditions with design or operating conditions. The accuracy of the model is important as it usually predicts reliability at tens or hundreds of years. There are many studies that had done on ALTS that used two stresses with different models, such as, E.A. Elsayed, Hao Zhang [3,4], Jung-Won Park and Bong-Jin Yum [5], Wenbiao Zhao [6] and Huan-Jyh Shyura, E.A. Elsayed and T.L James [7].
The most common models include:

1. 1). accelerated failure time models (AFT) where the failure times are inversely proportional to the applied stresses, i.e., failure times of the products at higher stresses are shorter than those tested at lower stresses [5].

2. 2). proportional hazards (PH) models where the failure rates are proportional to the applied stresses, i.e., the failure rates at higher stresses are higher than those at lower stresses [3].
3. 3). extended linear hazards regression models (ELHR) where both AFT and PH models are special cases of ELHR when some conditions are satisfied [7].
4. 4). proportional mean residual life models where the mean residual lives are inversely proportional to the stresses, i.e., the mean residual life at higher stresses is shorter than that at lower stresses [6];
5. 5). proportional odds (PO) models where the odds functions under different stress levels are proportional to each other [4].

2. Accelerated Failure Time models

The task of finding an ALT model can be divided into two steps:

1. 1). Choose an appropriate statistical distribution to describe lifetime at fixed levels of the accelerating variable(s). Typically the same distribution is used at all levels of stress.
2. 2). Choose a model to describe the relationship between the lifetime distributions and the accelerating variables.

If components are tested at different accelerated stress levels $s_1, s_2, \ldots, s_n$ The failure times at each stress level are then used to determine the most appropriate failure time probability distribution, along with its parameters. Under the accelerated failure time model assumptions, the failure times at different stress levels are linearly related to each other. Moreover, the failure time distribution at stress level s_1 is expected to be the same at different stress levels $s_2, s_3, \ldots, s_n$ as well as under the normal operating conditions. In other words, the shape parameters of the distributions are the same for all stress levels (including normal conditions) but the scale parameters may be different. Thus, the fundamental relationships between the operating conditions and stress conditions are summarized as follows:

1. Failure times:

$$t_0 = A_F \cdot t_s \tag{1}$$

   Where:
   t_o : is the failure time under operating conditions.
   t_s : is the failure time under stress conditions.
   A_F : is the acceleration factor (the ratio between product life under normal conditions and life under accelerated conditions).
2. Cumulative distribution functions ($CDFs$) :

$$F_0(t) = F_s\left(\frac{t}{A_F}\right) \tag{2}$$

3. Probability density functions:

$$f_0(t) = \left(\frac{1}{A_F}\right) \cdot f_s\left(\frac{t}{A_F}\right) \tag{3}$$

4. Failure rates:

$$h_0(t) = \left(\frac{1}{A_F}\right) \cdot h_s\left(\frac{t}{A_F}\right) \tag{4}$$

The accelerated failure time models are classified as parametric models and the most widely used models for the life distributions are the exponential and Weibull and lognormal models.

The life distribution is independent of applied stress so we can get the accelerated factor from life-stress relationship.
The most widely used basic relationships are:

1. 1). The Arrhenius relationship for temperature-accelerated tests.
2. 2). The inverse power relationship.

These relationships are for single applied stress, but because that our interest here is in utilization ALT with two stresses we will discuss in more details about these relationships. The generalized Eyring life-stress relationship is given by:

$$L(V,U) = \frac{1}{v} e^{A+\frac{B}{V}+CU+D\frac{U}{V}} \tag{5}$$

Where:

- V is the temperature in absolute units (such as Rankine or Kelvin).
- U is the non-thermal stress (i.e. voltage, vibration, etc.).
- A, B, C, D is the parameters to be determined.

The Eyring relationship is a special case of the generalized Eyring relationship where $C = D = 0$ and $A_{Eyr} = -A_{GEyr}$.
The generalized Eyring relationship includes the interaction of U and V as described by the $D\ fracUV$ term. In other words, this model can estimate the effect of changing one of the factors depending on the level of the other factor. The acceleration factor for the generalized Eyring relationship is given by:

$$A_F = \frac{L_{use}}{L_{accelerated}} = \frac{\frac{1}{V_U} e^{A+\frac{B}{V_U}+CU_U+D\frac{U_U}{V_U}}}{\frac{1}{V_A} e^{A+\frac{B}{V_A}+CU_A+D\frac{U_A}{V_A}}} \tag{6}$$

Where:

- L_{USE} is the life at use stress level.
- $L_{accelerated}$ is the life at the accelerated stress level.
- V_U is the use temperature level.
- V_A is the accelerated temperature level.

- U_A is the accelerated non-thermal level.
- U_U is the use non-thermal level.

There are two alternative life-stress relationships that do not consider interactions between stresses: the temperature-nonthermal life-stress relationship and the general log-linear relationship the two models are given by:
General Log-Linear:

$$L(U,V) = e^{\alpha_0 + \frac{\alpha_0}{V} + \alpha_1 U} \tag{7}$$

Temperature-NonThermal:

$$L(U,V) = \frac{c}{U^n e^{(\frac{B}{V})}} \tag{8}$$

There are another life-stress relationships that employ high temperature and humidity.Such as Pecks relationship:

$$\tau = A(RH)^{-n} exp[\frac{E}{KT}] \tag{9}$$

Intel (1988) uses another Eyring relationship:

$$\tau = Aexp(-B \cdot RH) exp[\frac{E}{KT}] \tag{10}$$

Intel notes that this differs little from Pecks relationship relative to uncertainties in such data.

3. Proportional Hazards PH Model

We assume the following model for the accelerated life test with two stresses:

1. 1). Two stresses z1 and z2 are used in the test, let z = (z1,z2) be the vector of stress levels.
2. 2). The proportional hazards model is employed to relate the reliability performance under different stress levels, it is expressed as:

$$\lambda(t;z) = \lambda_0(t) exp(\beta_1 Z_1 + \beta_2 Z_2) \tag{11}$$

   Where: β_1 and β_2 are unknown model parameters.
3. 3). The baseline hazard function $\lambda_0(t)$ is quadratic:

$$\lambda_0(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 \tag{12}$$

   Where: γ_0 , γ_1 and γ_2 are unknown parameters.

The hazard function $\lambda(t;z)$; is obtained by substituting $\lambda_0(t)$ into the PH model as:

$$\lambda(t;z) = (\gamma_0 + \gamma_1 t + \gamma_2 t^2) exp(\beta_1 Z_1 + \beta_2 Z_2) \tag{13}$$

We obtain the corresponding cumulative hazard function $\Lambda(t;z)$, reliability function $R(t;z)$ and density function $f(t;z)$ respectively, as follows:

$$\Lambda(t;z) = \int_0^t \lambda(u)\,du$$
$$\Lambda(t;z) = \int_0^t (\gamma_0 + \gamma_1 t + \gamma_2 t^2) exp(\beta_1 Z_1 + \beta_2 Z_2)\,du$$
$$\Lambda(t;z) = (\gamma_0 t + \frac{\gamma_1 t^2}{2} + \frac{\gamma_2 t^3}{3} exp(\beta_1 Z_1 + \beta_2 Z_2) \quad (14)$$
$$R(t;z) = exp(-\Lambda(t;z))$$
$$R(t;z) = exp\left(-(\gamma_0 t + \frac{\gamma_1 t^2}{2} + \frac{\gamma_2 t^3}{3} exp(\beta_1 Z_1 + \beta_2 Z_2)\right) \quad (15)$$
$$f(t;z) = \Lambda(t;z)R(t;z)$$
$$f(t;z) = (\gamma_0 + \gamma_1 t + \gamma_2 t^2) exp(\beta_1 Z_1 + \beta_2 Z_2) \quad (16)$$
$$\cdot \exp\left(-(\gamma_0 t + \frac{\gamma_1 t^2}{2} + \frac{\gamma_2 t^3}{3}) exp(\beta_1 Z_1 + \beta_2 Z_2)\right) \quad (17)$$

In order to obtain the values of the model parameters, we utilize the maximum partial likelihood estimation (MLE) procedure [3].

4. Extended hazard Regression Model

Let $\Lambda(t;z)$ denote the hazard rate at time t for a component with covariate vector z(stress z). The vector z is a set of applied stresses accounting for the effects of environmental stresses on the hazard rate. The basic extended hazard regression model is defined as follows:

$$\lambda(t;z) = g_1(\alpha \cdot z) \cdot \lambda_0(g_2(\beta \cdot z)t) \quad (18)$$

If we assume that $g_1(x) = g_2(x) = exp(x)$ then the EHR model is:

$$\lambda(t;z) = e^{\alpha z} \cdot \lambda_0(e^{\beta\ cdotz} \cdot t) \quad (19)$$

Where:
$\alpha z = \alpha_1 z_1 + \alpha_2 z_2 + \cdots + \alpha_k z_k$
and $\beta z = \beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_k z_k$

Based on the EHR model, we generate the corresponding mathematical functions of reliability and mean time to failure ($MTTF$). The reliability function

is:

$$\Lambda(t;z) = e^{\alpha z} \cdot \lambda_0(e^{\beta \cdot z} \cdot t) \tag{20}$$

$$R(x;z) = exp\left(-\int_0^t \lambda(x;z)\,dx\right) \tag{21}$$

$$ln(R(x;z)) = -\int_0^t \lambda(x;z)\,dx$$

$$ln(R(x;z)) = -\int_0^t e^{\alpha z} \cdot \lambda_0(e^{\beta\ cdotz} \cdot x)\,dx$$

$$ln(R(x;z)) = e^{\alpha z}\left[-\int_0^t \lambda_0(e^{\beta\ cdotz} \cdot x)\,dx\right]$$

$$ln(R(x;z)) = \frac{e^{\alpha z}}{e^{\beta z}}\left[-\int_0^{\zeta} \lambda_0(u)\,du\right]$$

$$ln(R(x;z)) = \frac{e^{\alpha z}}{e^{\beta z}} ln\left[exp\left(-\int_0^{\zeta} \lambda_0(u)\,du\right)\right]$$

$$ln(R(x;z)) = \frac{e^{\alpha z}}{e^{\beta z}} ln\left[S_0(\zeta)\right] \tag{22}$$

Where :

$\zeta = e^{\beta z} \cdot t$

$S_0(\zeta) = exp\left(-\int_0^{\zeta} \lambda_0(u)\,du\right)$

The baseline hazard function λ_0 is specified by a quadratic spline function to estimate the unknown underlying distribution. The formula is given by:

$$\lambda_0(u) = \sum_{n=0}^{2} \gamma_n \mu^n + \sum_{m=1}^{l} \theta_m (\mu - \tau_m)^2 \tag{23}$$

Where:

l is the number of knots.

τ_m is the location of knots.

θ_m is the added linear effect following of knots.

And γ_n are the coefficients of the underlying base polynomial (Splines are presented as a nonparametric function estimating technique. A spline function of degree m is a piecewise m-degree polynomial with pieces joining at defined points, which are called knots. [8]). The unknown parameters of the EHR model are estimated by use the maximum likelihood approach [7].

5. Proportional Mean Residual Life Regression Model

Two distributions with reliability functions R_0 and R with mean residual lives at time x of $e_0(x)$ and $e(x)$ respectively, are said to have proportional MRL functions,

if they are related as follows.

$$e(x) = \theta e_0(x) \qquad \forall x, \theta > 0 \tag{24}$$

Therefore:

$$R(x,\theta) = \frac{e(0)}{e(x)} exp\left(-\int_0^x \frac{du}{e(u)}\right)$$
$$R(x,\theta) = R_0(x)\left[\frac{\int_0^x}{\infty}\frac{R_0(u)du}{\mu_0}\right]^{\frac{1}{\theta}-1} \tag{25}$$

where: $\mu_0 = e_0(0)$
Thus the probability density function is:

$$f(t) = \frac{e_0(0)}{e_0^2(t)} exp\left(-\int_0^t \frac{du}{\theta e_0(u)}\right)\left(\frac{1}{\theta} + e_0^{'}(t)\right) \tag{26}$$

We extend the model to a more general framework with a vector Z .

$$e(t \mid z) = exp(\beta^T \cdot z) e_0(t) \tag{27}$$

We refer to this model as the proportional mean residual life regression model which we utilize to model accelerated life testing. Clearly $e_0(t)$ serves as the MRL corresponding to a baseline reliability function $R_0(x)$ and $e(t \mid z)$ is the conditional mean residual life function and Z Covariates (applied stresses) or explanatory variables are of primary importance in accelerated life testing. It is of a great interest to model the effect of covariates on failure times. For instance, temperature and electric field are some of the covariates of interest.

$Z^T = (Z_1, Z_2, \ldots, Z_p)$ and $\beta^T = (\beta_1, \beta_2, \ldots, \beta_p)$

Typically, we obtain experimental data in the form $((t_i, z_i); i = 1, \ldots, n)$ which represents the set of failure times and the vectors of covariates for each unit. The PMRL model is a nonparametric multiple regression approach for reliability estimation, in which the baseline MRL function is modified multiplicatively by the covariates. If we assume the baseline MRL $e_0(t)$ to be exponential with time:

$$e_0(t) = \gamma_0 exp(\gamma_1 t) \tag{28}$$

Substituting $e_0(t)$ into the PMRL model, we obtain:

$$e(t; z) = \gamma_0 exp(\beta^T \cdot z + \gamma_1 t) \tag{29}$$

We obtain the corresponding hazard function, and the reliability function $R(t; z)$

as:

$$\lambda(t;z) = \frac{e_0'(t) + exp(-\beta^T \cdot z)}{e_0(t)}$$

$$\lambda(t;z) = \frac{\gamma_0\gamma_1 exp(\gamma_1 t) + exp(-\beta^T \cdot z)}{\gamma_0 exp(\gamma_1 t)} \tag{30}$$

$$R(t;z) = \frac{e_0(0)}{e_0(t)} exp\left(-\int_0^t \frac{du}{\theta e_0(u)}\right)$$

$$R(t;z) = exp(-\gamma_1 t) exp\left(-exp(-\beta^T \cdot z) - \int_0^t \frac{exp(-\gamma_1 t)du}{\gamma_0}\right)$$

$$R(t;z) = exp[-\gamma_1 t + \frac{exp(-\beta^T \cdot z)(exp(-\gamma_1 t) - 1)}{\gamma_0\gamma_1}] \tag{31}$$

The pdf is obtained as:

$$f(t;z) = \frac{e_0(0)}{e_0^2(t)} exp\left(-\int_0^t \frac{du}{\theta e_0(u)}\right)\left(\frac{1}{\theta} e_0'(t)\right)$$

$$f(t;z) = \frac{exp(-2\gamma_1 t)}{\gamma_0} exp\left(-exp(-\beta^T \cdot z) \cdot \int_0^t \frac{exp(-\gamma_1 t)du}{\gamma_0}\right) \tag{32}$$

$$\cdot (exp(-\beta^T \cdot z) + \gamma_0\gamma_1 exp(\gamma_1 t))$$

$$f(t;z) = \frac{1}{\gamma_0} exp\left[-2\gamma_1 t + \frac{exp(-\beta^T \cdot z)(exp(-\gamma_1 t) - 1)}{\gamma_0\gamma_1}\right] \tag{33}$$

$$\cdot (exp(-\beta^T \cdot z) + \gamma_0\gamma_1 exp(\gamma_1 t)) \tag{34}$$

With the maximum likelihood method we can estimates the parameter γ_0,γ_1 and β of the model [6].

6. Proportional odds PO Model

This model is referred to Proportional Odds (PO) model since the odds ratios under different stress levels are proportional to each other as:

$$\frac{F(t)}{1 - F(t)} = e^{\beta z} \frac{F_0(t)}{1 - F_0(t)} \tag{35}$$

Assuming two stresses exist, after mathematical manipulation, last equation could be expressed as:

$$\lambda(t) = \frac{e^{\beta_1 z_1 + \beta_2 z_2} \lambda_0(t)}{1 - (1 - e^{\beta_1 z_1 + \beta_2 z_2}) F_0(t)} \tag{36}$$

The baseline hazard function $\lambda_0(t)$ is assumed to be quadratic:

$$\lambda_0(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 \tag{37}$$

So we rewrite $\lambda(t)$ as:

$$\lambda(t) = \frac{e^{\beta_1 z_1+\beta_2 z_2}(\gamma_0 + \gamma_1 t + \gamma_2 t^2)}{e^{\beta_1 z_1+\beta_2 z_2} + (1 - e^{\beta_1 z_1+\beta_2 z_2})e^{-(\gamma_0 t+\gamma_1 t^2/2+\gamma_2 t^3/2)}} \tag{38}$$

We also have:

$$F_0(t) = 1 - e^{-\int_0^t \lambda_0(u)du} = 1 - e^{-(\gamma_0 t+\gamma_1 t^2/2+\gamma_2 t^3/2)} \tag{39}$$

$$R_0(t) = exp[-\Lambda_0(t)] = 1 - F_0(t) = e^{-(\gamma_0 t+\gamma_1 t^2/2+\gamma_2 t^3/2)}] \tag{40}$$

$$\Lambda_0(t) = \int_0^t \lambda_0(u)\, du = \gamma_0 t + \gamma_1 t^2/2 + \gamma_2 t^3/2 \tag{41}$$

$$\Lambda(t; z) = \int_0^t \lambda(u; z)\, du = \int_0^t \frac{e^{\beta z}\lambda_0(u)}{e^{\beta z} + (1 - e^{\beta z})R_0(u)}\, du$$

$$\Lambda(t; z) = \int_0^t \frac{e^{\beta_1 z_1+\beta_2 z_2}(\gamma_0 + \gamma_1 t + \gamma_2 t^2)}{e^{\beta_1 z_1+\beta_2 z_2} + (1 - e^{\beta_1 z_1+\beta_2 z_2})e^{-(\gamma_0 t+\gamma_1 t^2/2+\gamma_2 t^3/2)}} \tag{42}$$

The values of γ_0 , γ_1 , γ_2 , β_1 and β_2 can be estimated by maximizing log likelihood function which are obtained by numerical method [3].

7. Conclusion

In the accelerated failure time models components are tested at different stress levels and the failure times are then used to determine the most appropriate failure time distribution and its parameters. In this case the failure times follow the same general distributions for all different stress levels, including the normal operating conditions. But When the failure time data involve a complex lifetime distribution or when the number of observations is small, making it difficult to fit the failure time distribution accurately, the rest models (proportional hazards (PH), extended linear hazards regression ELHR, proportional mean residual life PMRL, proportional odds (PO) models) which are classified as non-parametric models appear to be a very attractive approach to predict reliability at different stress levels.The advantage of these models is that they are essentially "distribution free".

References

1. Wayne Nelson, *Accelerated testing: Statistical models, test plans, and data analyses*, New York, Wiley, 1990.
2. Meeker W.Q., Escobar L.A., *Statistical Methods for Reliability Data*, AP&S, Wiley,1998
3. E.A. Elsayed, H. Zhang, *Design of PH-based accelerated life testing plans under multiple-stress-type*, Reliability Engineering and System Safety, 2007

4. E.A. Elsayed, H. Zhang, *Design of optimum reliability test plans under multiple stresses*, QUALITA 2005, quality and dependability, Bordeaux, France, March 1618, 2005.
5. Park JW, Yum BJ, *Optimum design of accelerated life tests with two stresses*, Nav Res Logistic, 1996.
6. Wenbiao Zhao, E.A. Elsayed , *Optimum Accelerated Life Testing Plans Based Proportional Mean Residual Life*, Reliability Engineering and System Safety, 2007
7. Huan-Jyh Shyura, E.A. Elsayedb and James T. Luxhj, *A general hazard regression model for accelerated life testing*, Annals of Operations Research, 1999.
8. E.J. Wegmen and I.W. Wright, *Splines in statistics*, Journal of the American Statistical Association, 1983.
9. Hoang Pham Editor, *Handbook of Reliability Engineering*, Springer-Verlag London Limited, 1 edition , April 15, 2003.

An Approach For The Equality Of Coefficients Of Variation For Normal Populations

M. R. Kazemi, A. A. Jafari and J. Behboodian
Islamic Azad University-Kazerun Branch

Department of Statistics, Yazd University
Department of Statistics, Shiraz University

Department of Mathematics, Shiraz Islamic Azad University

In this article, a new approach is given for the equality of coefficient of variation of k normal populations. This approach is based on parametric bootstrap method. Simulation studies show that the size of this test is smaller than nominal level and the new approach is better than other existing methods. Two real examples are proposed for illustrating our new approach.

Keywords: Coefficient of Variation, Monte Carlo simulation, Parametric bootstrap

1. Introduction

Coefficient of variation (CV) is an important unitless of measure of relative variability and widely used in many practical applications such as biological and financial sciences, diagnostic areas, medicine, and clinical experiments. It is defined as the ratio of mean to standard deviation and populations can have a same CV even if the means and variances are different. If the variances of the populations are equal or known then we can use ANOVA or a similar ANOVA test for equality of CV's of populations. But in applications variances are unknown and there is not any information about them.

Many tests have been proposed for the equality of CV's of k normal populations; Bennett (1976) presented a likelihood ratio test using Makay approximation. Doornbos and Dijkstra (1983) developed a noncentral t test. Shafer and Sullivan (1986) presented a modified version of Bennett's test. Miller (1991b) provided a nonparametric square rank test. Miller (1991a) and Feltz and Miller (1996) derived one, two and k-sample tests for CV of normal populations. Rao and Vidya (1992), Gupta and Ma (1996), and Rao and Jose (2001) provided Wald tests for the equality of two and more than two populations. Pardo and Pardo (2000) obtained

a test based on Renyi's divergence. Nairy and Rao (2003) developed likelihood ratio test, Wald test and score test for the equality of inverse CV's. Verrill and Johnson (2007a) provided a likelihood ratio test based on one step Newton estimators. Tsou (2009) obtained a robust score test for the equality of CV's of k normal and nonnormal populations.

Simulations studies are performed by Feltz and Miller (1996), Gupta and Ma (1996), Fung and Tsang (1998), Pardo and Pardo (2000), Nairy and Rao (2003).

The aim of this article is to develop a parametric bootstrap (BP) approach for testing the equality of several normal populations. Bootstrap is used frequently in applied statistics. In fact, it is a computer method that applied on observed data by Monte Carlo Simulation (Efron and Tibshirani, 1993). When the data have a certain model, this method is called parametric bootstrap and usually has higher efficiency than when the model is unknown (nonparametric bootstrap). Krishnamoorthy et al. (2007) proposed a PB approach for testing the equality of means of several normal populations. We used this idea for comparing the CV's of k normal populations.

This paper is organized as follows; In Section 2, a parametric bootstrap test is developed for testing the equality of CV's and a computational Algorithm is described, and also, we review some other approximation methods, briefly. A simulation study is performed in Section 3 for comparing the sizes of approaches that are given is Section 2. Two real examples are illustrated in section 4.

2. Approaches for the equality of CV's

Let X_{ij}, $i = 1, ..., k$, $j = 1, ..., n_i$, be k independent random samples of size n_i from normal populations with $E(X_{ij}) = \mu_i$, and $Var(X_{ij}) = \sigma_i^2$, i.e. $X_{ij} \sim N(\mu_i, \sigma_i^2)$. The CV for the ith population is defined as $\varphi_i = \dfrac{\sigma_i}{\mu_i}$, $i = 1, ..., k$. The hypotheses of interest are

$$H_0 : \varphi_1 = ... = \varphi_k = \varphi \quad vs. \quad H_1 : \varphi_i \neq \varphi_j, \text{ for some } i \neq j \tag{1}$$

where φ is the unknown common CV parameter.

Denote the sample mean and sample variance of the ith population as $\bar{X}_i$, and S_i^2, respectively, where $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ and $S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2$.

Consider $\theta_i = \dfrac{1}{\varphi_i} = \dfrac{\mu_i}{\sigma_i}$ is the inverse CV for ith population. Then the hypotheses in (1) are equivalent to

$$H_0 : \theta_1 = ... = \theta_k = \theta \quad vs. \quad H_1 : \theta_i \neq \theta_j, \text{ for some } i \neq j \tag{2}$$

where θ is unknown.

We know that $\frac{\bar{X}_i}{\sigma_i} \sim N(\theta_i, \frac{1}{n_i})$. Therefore

$$\mathbf{Hy} \sim N_{k-1}(\mathbf{H}\underset{\sim}{\theta}, \mathbf{HVH}'),$$

where $\mathbf{y} = (\frac{\bar{X}_1}{\sigma_1}, ..., \frac{\bar{X}_k}{\sigma_k})$, $\mathbf{V} = [diag(\frac{1}{n_1}, ..., \frac{1}{n_k})]$, $\underset{\sim}{\theta} = (\theta_1, ..., \theta_k)$, $\mathbf{H} = [\mathbf{1} : \mathbf{B}]$, $\mathbf{1} = (1, ..., 1)'$, and $\mathbf{B} = [diag(-1, ..., -1)]$.

If σ_i^2's are known, then a natural statistic for testing (2) is given by

$$Q(\bar{X}_i, ..., \bar{X}_k; \sigma_1, ..., \sigma_k) = \sum_{i=1}^{k} n_i \left[\frac{\bar{X}_i}{\sigma_i} - \hat{\theta}\right]^2 = \sum_{i=1}^{k} \frac{n_i \bar{X}_i^2}{\sigma_i^2} - \frac{(\sum_{i=1}^{k} n_i \bar{X}_i/\sigma_i)^2}{n}$$
$$= (\mathbf{Hy})'(\mathbf{HVH}')^{-1}(\mathbf{Hy}), \qquad (3)$$

where $n = \sum_{i=1}^{k} n_i$, and $\hat{\theta} = \frac{\sum_{i=1}^{k} n_i \bar{X}_i/\sigma_i}{n}$ is Maximum likelihood estimation (MLE) for θ.

Therefore, $Q(\bar{X}_1, ..., \bar{X}_k; \sigma_1, ..., \sigma_k)$ has a noncentral chi-square distribution with $k-1$ degrees of freedom and noncentrality parameter λ. It can be easily shown that the noncentrality is

$$\lambda = (\mathbf{H}\underset{\sim}{\theta})'(\mathbf{HVH}')^{-1}(\mathbf{H}\underset{\sim}{\theta}) = \sum_{i=1}^{k} n_i \theta_i^2 - \frac{(\sum_{i=1}^{k} n_i \theta_i)^2}{n}.$$

If null hypothesis in (2) occurs then $\lambda = 0$. Thus $Q(\bar{X}_i, ..., \bar{X}_k; \sigma_1, ..., \sigma_k)$ rejects H_0 in level α iff

$$Q(\bar{X}_1, ..., \bar{X}_k; \sigma_1, ..., \sigma_k) > \chi^2_{k-1,\alpha},$$

where $\chi^2_{k-1,\alpha}$ is the $(1-\alpha)$th quantile of a chi-square distribution with $k-1$ degrees of freedom.

In general, the σ_i^2's are unknown; in this case by replacing σ_i in (3) by S_i, we obtain a test statistic that is given by

$$Q(\bar{X}_1, ..., \bar{X}_k; S_1, ..., S_k) = \sum_{i=1}^{k} \frac{n_i \bar{X}_i^2}{S_i^2} - \frac{(\sum_{i=1}^{k} n_i \bar{X}_i/S_i)^2}{n}$$
$$= \sum_{i=1}^{k} n_i \hat{\theta}_i^{*2} - \frac{(\sum_{i=1}^{k} n_i \hat{\theta}_i^*)^2}{n}, \qquad (4)$$

where $\hat{\theta}_i^* = \frac{\bar{X}_i}{S_i}$, and for large n, $Q(\bar{X}_1, ..., \bar{X}_k; S_1, ..., S_k)$ has a chi-square distribution with $k-1$ degrees of freedom.

2.1. *Parametric bootstrap*

In this section, we develop a parametric method for testing the equality of CV's. This parametric bootstrap approach involves sampling from estimated models and the parametric bootstrap pivotal variable can be developed as follows.

The log-likelihood function of $X_{ij},\ i = 1, ..., k,\ j = 1, ..., n_i$ under the null hypothesis in (2) can be written as

$$\ell(\sigma_1, ..., \sigma_k, \theta) = -\frac{1}{2}\sum_{i=1}^{k} n_i ln 2\pi\sigma_i^2 - \sum_{i=1}^{k}\frac{1}{2\sigma_i^2}((n_i-1)s_i^2 + n_i(\bar{x}_i - \theta\sigma_i)^2).$$

Let $(\tilde{\sigma}_1, ..., \tilde{\sigma}_k, \tilde{\theta})$ be the MLE of $(\sigma_1, ..., \sigma_k, \theta)$, which can be calculated using the numerical methods.

Let $\bar{X}_{Bi} \sim N(\tilde{\theta}, \frac{1}{n_i})$ and $S_{Bi}^2 \sim \chi_{n_i-1}^2/(n_i-1), i = 1, ..., k$. Then the parametric bootstrap pivotal variable based on the test statistic in (4) is given by

$$Q(\bar{X}_{B1}, ..., \bar{X}_{Bk}; S_{B1}, ..., S_{Bk}) = \sum_{i=1}^{k}\frac{n_i\bar{X}_{Bi}^2}{S_{Bi}^2} - \frac{(\sum_{i=1}^{k} n_i\bar{X}_{Bi}/S_{Bi})^2}{n}. \qquad (5)$$

Note that $\bar{X}_{Bi}$ is distributed as $\frac{1}{\sqrt{n_i}}Z_i + \tilde{\theta}$, where Z_i is a standard normal random variable. Therefore the parametric bootstrap pivotal variable in (5) is distributed as

$$Q_B = \sum_{i=1}^{k}\frac{n_i(\frac{1}{\sqrt{n_i}}Z_i + \tilde{\theta})^2}{\chi_{n_i-1}^2/(n_i-1)} - \frac{(\sum_{i=1}^{k} n_i\sqrt{n_i-1}(\frac{1}{\sqrt{n_i}}Z_i + \tilde{\theta})/\sqrt{\chi_{n_i-1}^2})^2}{n}. \qquad (6)$$

For given $\tilde{\theta}$, MLE of θ, the parametric bootstrap test rejects the null hypothesis in (2) at level α when

$$P(Q_B \geq Q_{B0}) \leq \alpha \qquad (7)$$

where Q_{B0} is observed value of $Q(\bar{X}_1, ..., \bar{X}_k; S_1, ..., S_k)$ in (4).

The probability in (7) can be estimated using the Monte Carlo simulation given in following Algorithm.

Algorithm 2.1. *For given* $(n_1, ..., n_k)$, $(\bar{x}_1, ..., \bar{x}_k)$, *and* $(s_1^2, ..., s_k^2)$,
compute Q_{B0} *in (4).*
calculate $\tilde{\theta}$, *MLE of* θ.
For $j = 1, ..., M$
generate $Z_i \sim N(0, 1)$ *and* $\chi_{n_i-1}^2$, $i = 1, ..., k$.
compute Q_B *using (6).*
Let $W_j = 1$ *if* $Q_B > Q_{B0}$, *else* $W_j = 0$.

(end loop)
$\frac{1}{M}\sum_{j=1}^{M} W_j$ *is a Monte Carlo estimate of p-value in (7).*

2.2. *Bennet's test*

Consider the following test statistic

$$BT = (n-k)log\left(\frac{\sum_{i=1}^{k} d_i^*}{n-k}\right) - \sum_{i=1}^{k}(n_i-1)log\left(\frac{d_i^*}{n_i-1}\right), \tag{8}$$

where $d_i^* = \frac{n_i\hat{\varphi}_i^{*2}}{\hat{\varphi}_i^{*2}+1}$, $\hat{\varphi}_i^* = \frac{S_i}{\bar{X}_i}$, and $S_i^2 = \frac{1}{n_i-1}\sum_{i=1}^{k}(X_{ij}-\bar{X}_i)^2$.

BT under the null hypothesis has a chi-square distribution with $k-1$ degrees of freedom and rejects H_0 if BT$> \chi^2_{\alpha,(n-1)}$.

2.3. *Modified Bennet's test*

Shafer and Sullivan (1986) consider the modified Bennet's test statistic as

$$MBT = (n-k)log\left(\frac{\sum_{i=1}^{k} d_i}{n-k}\right) - \sum_{i=1}^{k}(n_i-1)log\left(\frac{d_i}{n_i-1}\right), \tag{9}$$

where $d_i = \frac{n_i\hat{\varphi}_i^{2}}{\hat{\varphi}_i^{2}+1}$, $\hat{\varphi}_i = \frac{S_{i(b)}}{\bar{X}_i}$, and $S_{i(b)}^2 = \frac{1}{n_i}\sum_{i=1}^{k}(X_{ij}-\bar{X}_i)^2$.

MBT under the null hypothesis also has a chi-square distribution with $k-1$ degrees of freedom and rejects H_0 if MBT$> \chi^2_{\alpha,(n-1)}$.

2.4. *Wald test*

Nairy and Rao (2003) consider a Wald test for equality of inverse CV's as the following statistic

$$WT = \hat{h}_1'\left[H\hat{V}H'\right]^{-1}\hat{h}_1, \tag{10}$$

where $h = (\theta_1-\theta_2, ..., \theta_1-\theta_k)$, $\theta_i = \frac{\mu_i}{\sigma_i}$, $\hat{V} = \left[diag(\frac{2+\hat{\theta}_1^2}{2n_1}, ..., \frac{2+\hat{\theta}_k^2}{2n_k})\right]$, H is the matrix in (3.3), and $\hat{\theta}_i = \frac{\bar{X}_i}{S_{i(b)}}$. Easily, we can show that

$$WT = \sum_{i=1}^{k}\frac{2n_i\hat{\theta}_i^2}{2+\hat{\theta}_i^2} - \frac{\left(\sum_{i=1}^{k}\frac{2n_i\hat{\theta}_i}{2+\hat{\theta}_i^2}\right)^2}{\sum_{i=1}^{k}\frac{2n_i}{2+\hat{\theta}_i^2}}. \tag{11}$$

WT under the null hypothesis has a chi-square distribution with $k-1$ degrees of freedom and rejects H_0 if BT$> \chi^2_{\alpha,(n-1)}$.

2.5. *Modified Miller test*

Feltz and Miller (1996) suggested a modified Miller test. Their proposed test statistic

$$MT = \varphi^{-2}(0.5+\varphi^2)^{-1}\left[\sum_{i=1}^{k}(n_i-1)\left(\hat{\varphi}_i^* - \frac{\sum_{i=1}^{k}(n_i-1)\hat{\varphi}_i^*}{n-k}\right)\right], \qquad (12)$$

where $\hat{\varphi}_i^* = \dfrac{S_i}{\bar{X}_i}$, is asymptotically chi-square distributed with $k-1$ degrees of freedom under the null hypothesis. If φ is replaced by $\dfrac{\sum_{i=1}^{k}(n_i-1)\hat{\varphi}_i^*}{n-k}$, the asymptotic distribution of statistic is unaffected.

3. Simulation study

A Monte Carlo simulation is performed for comparing the estimated type I error probabilities of the given approaches in Section 2. n_i, $i = 1, ..., k$ observations were generated from k normal populations with mean 100 and standard deviation $10a$, $a = 1, ..., 5$. In fact, all populations have a same CV, $0.1a$. In practice, CV rarely exceed 0.5 for medical and biological sciences (Fung and Tsang, 1998). In this simulation we we consider $\alpha = 0.05$, and used 10000 times runs for each of the sample size and parameter configurations.

In Tables 1, we only present the results of simulations for $a = 1$, i.e. when the CV's of populations equal 0.1. For other cases we found similar results and did not give here. It can be found that the type I error probabilities of PB are always smaller than nominal level, $\alpha = 0.05$, but size of other approaches are greater than nominal level, especially when the sample sizes are small.

Table 1. Estimated type I error probabilities for tests at $\alpha = 0.05$.

	Test				
$(n_1, ..., n_k)$	PB	BT	MBT	WT	MT
(3,3,3,3,3)	0.030	0.087	0.089	0.077	0.069
(3,4,5,6,7)	0.035	0.074	0.073	0.066	0.058
(3,3,3,5,5)	0.030	0.076	0.075	0.064	0.055
(3,10,10,10,10)	0.034	0.073	0.072	0.066	0.060
(3,20,20,20,20)	0.047	0.068	0.062	0.058	0.050
(5,5,5,5,5)	0.029	0.066	0.070	0.063	0.060
(5,5,5,10,30)	0.040	0.068	0.061	0.057	0.044
(5,5,5,30,30)	0.043	0.072	0.063	0.060	0.044
(5,6,7,8,9)	0.034	0.065	0.065	0.062	0.058
(5,10,15,20,25)	0.046	0.062	0.059	0.056	0.050
(5,10,10,20,20)	0.040	0.063	0.061	0.059	0.054
(5,30,30,30,30)	0.046	0.060	0.058	0.057	0.052
(7,7,7,7,7)	0.031	0.061	0.061	0.058	0.056
(7,8,9,10,11)	0.035	0.059	0.060	0.058	0.054
(7,7,7,30,30)	0.044	0.059	0.054	0.055	0.045
(7,30,30,30,30)	0.044	0.056	0.055	0.055	0.052
(10,10,10,10,10)	0.036	0.058	0.058	0.060	0.058
(10,11,12,13,14)	0.037	0.056	0.056	0.056	0.054
(10,10,10,20,20)	0.039	0.057	0.057	0.056	0.053
(10,20,20,30,30)	0.042	0.053	0.053	0.054	0.052
(20,20,20,20,20)	0.039	0.052	0.052	0.054	0.053
(20,21,21,23,24)	0.041	0.055	0.055	0.055	0.055
(20,20,20,30,30)	0.038	0.053	0.051	0.052	0.052
(20,30,30,30,30)	0.044	0.056	0.057	0.056	0.055
(30,30,30,30,30)	0.042	0.052	0.052	0.051	0.051

4. Real Data Analysis

We shall illustrate the introduced methods in Section 2 for testing the equality of CV's of populations using two real examples; (i) Parametric bootstrap (PB) (ii) Bennet's test (BT) (iii) Modified Bennet's test (MBT) (iv) Wald test (WT) (v) Modified Miller test (MT). These two real examples are proposed by Nairy and Rao (2003). We note that the p-value for PB is obtained using Algorithm 1 with $M = 100000$.

Example 4.1. Over nine years from 1991-1999 for the State of Karnataka, India, this data is collected to catches of four kinds of fish . For these four kinds of fish the estimate of CV's, $\bar{x}_i$, $i = 1, 2, 3, 4$, are 605.67, 109.00, 303.00, and 93.44, respectively. Also s_i^2, $i = 1, 2, 3, 4$, 44460.00, 13313.50, 12983.25, and 6886.78, respectively. The p-values for PB, BT, MBT, WT, and MT are 0.0809, 0.0534, 0.0439, 0.0301, and 0.0609, respectively

Example 4.2. This data refers to survival of patients from four hospital, which is a part of the data in Appendix D of the Fleming and Harrinton (1991). The sample size for these hospitals are 5, 4, 3, and 10. $\bar{x}_i$, $i = 1, 2, 3, 4$, are 168.00, 59.50, 45.67, and 155.50, respectively. Also s_i^2, $i = 1, 2, 3, 4$, are 6880.50, 4460.33, 714.33, and 9105.61, respectively. The p-values for PB, BT, MBT, WT, and MT are 0.8353, 0.7064, 0.7064, 0.6277, and 0.6051, respectively.

References

1. Bennett, B. M. (1976). On an approximate test for homogeneity of coefficients of variation, in Ziegler, W.J. (ed.), *Contribution to Applied Statistics*, Birkhauser Verlag, Basel and Stuttgart, 169-171.
2. Doornbos, R. and Dijkstra, J. B. (1983). A multi sample test for the equality of coefficients of variation in normal populations, *Commun. Statist. Simul. Computat.*, 12, 147-158.
3. Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. London: Chapman & Hall.
4. Feltz, C. J. and Miller, G. E.(1996). An asymptotic test for the equality of coeffcients of variation from k populations, *Statist. Med.*, 15, 647-658.
5. Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*, New York; John Wiley.
6. Fung, W. K. and Tsang, T. S. (1998). A simulation study comparing tests for the equality of coefficients of variation, *Statist. Med.*, 17, 2003-2014
7. Gupta, R. C. and Ma, S. (1996). Testing the equality of coefficients of variation in k normal populations, *Commun. Statist. Theor. Meth*, 25, 115-132.
8. Krishnamoorthya, K., Lu, F. and Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: fixed and random models, *Computational Statistics & Data Analysis*, 51, 5731-5742.
9. Miller, G. E. (1991a). Asymptotic test statistics for coefficients of variation, *Commun. Statist. Theor. Meth.*, 20, 3351-3363.
10. Miller, G. E. (1991b). Use of the squared ranks test to test for the equality of the coefficients of variation, *Commun. Statist. Simul. Computat.* 20, 743-750.
11. Nairy, K. S. and Rao, K. A. (2003). Tests of coefficient of variatin of normal population, *Comm. Stat., Simul. Computat.,* 32, 641-661.
12. Pardo, M. C. and Pardo, J. A. (2000). Use of Rényi's divergence to test for the equality of the coefficient of variation, *J. Comp. Appl. Math.*, 116, 93-104.
13. Rao, K. A. and Jose, C. T. (2001). Test for Equality of Coefficient of Variation of K Populations, In: *Proceeding the 53^{rt} Session of International Statistical Institude*, Seoul, Korea.
14. Rao, K. A. and Vidya, R. (1992). On the performance of test for coefficient of variation, *Calcutta Statistical Association Bulletin*, 42, 87-95.
15. Shafer, N. J. and Sullivan, J. A. (1986). A simulation study of a test for the equality of the coefficients of variation, *Comm. Stat., Simul. Computat.*, 15, 681-695.
16. Tsou. T. S (2009). A Robust Score Test for Testing Several Coefficients of Variation with Unknown Underlying Distributions, *Commun. Statist. Theor. Meth.*, 38, 1350-1360.

A Note on The Mean Residual Life Function of a Coherent System With Exchangeable or Nonidentical Components

M. Khanjari Sadegh (Invited)

Department of Statistics,University of Birjand

This article investigates some properties of the mean residual life function of a coherent system consisting of n components in two different cases, when the lifetimes of the system components are independent random variables but not necessary identically distributed and when the joint distribution of the components lifetimes is exchangeable. Our results are mainly about the two following mean residual life functions of a $(n-k+1)-$out-of-n system

$$H_n^k(t) = E(T_{k:n} - t | T_{1:n} > t),$$
$$M_n^{r,k}(t) = E(T_{k:n} - t | T_{r:n} > t), \quad 1 \le r \le k \le n$$

where $T_{1:n}, T_{2:n}, \ldots, T_{n:n}$ are the ordered statistics corresponding to $T_1, T_2, \ldots, T_n$ the lifetimes of the system components. We obtain a sufficient condition for $H_n^k(t)$ to be a decreasing(increasing) function of t, in nonidentical case and for $H_n^1(t)$ in exchangeable case. We have shown in nonidentical case that when the components of the system have increasing failure rate, $M_n^{r,k}(t)$ is decreasing in time. In both above mentioned cases it is shown that $M_n^{k,k}(t) \le M_n^{r,k}(t) \le H_n^k(t)$. Also it is shown in both cases that $M_n^{r,k}(t)$ is a decreasing function of r. We show that $H_{n-1}^{k-1}(t) \le H_n^k(t) \le H_{n-1}^k(t)$, if T_n is independent of $T_1, \ldots, T_{n-1}$. Finally using properties of Samaniego's signature, the extension to a general coherent system with exchangeable components is given. Our results extend the results in Khanjari (2008 Comm. in Statistics) for the case of a parallel system with nonidentical components and the results in Asadi and Goliforushani (2008, IEEE Trans. on Rel.) for the case of a coherent system with independent and identical components.

Keywords: Mean residual life function, Order statistics, $(n-k+1)$-out-of-n system, Nonidentical components, Exchangeable components, Failure rate, IFR, DFR, Signature

1. Introduction

Consider a coherent system consisting of n components and let $T_1, T_2, \ldots, T_n$ denote the lifetimes of the system components which are nonnegative and absolutely continuous random variables. We first assume that T_i's are independent but not identical and then assume that T_i's have an exchangeable joint distribution. We also use $T_{1:n}, T_{2:n}, \ldots, T_{n:n}$ to represent the ordered lifetimes of the components. It is well known that the lifetime of a $(n-k+1)$-out-of-n system is $T_{k:n}$. The mean residual life (MRL) function of this system is considered in Section 2. Under two different described cases for components lifetimes, some properties of

$H_n^k(t) = E(T_{k:n} - t | T_{1:n} > t)$ and $M_n^{r,k}(t) = E(T_{k:n} - t | T_{r:n} > t)$, $1 \le r \le k \le n$ are obtained. $H_n^k(t)$ is the MRL function of a $(n-k+1)$-out-of-n system in which, at time t, all components of the system are operating but in $M_n^{r,k}(t)$ at least $n-r+1$ components of the system are working at the same time t. We obtain a sufficient condition for $H_n^k(t)$ to be a decreasing(increasing) function of t, in nonidentical case and for $H_n^1(t)$ in exchangeable case. We have shown in nonidentical case that when the components of the system have increasing failure rate, $M_n^{r,k}(t)$ is decreasing in time. In both above mentioned cases it is shown that $M_n^{k,k}(t) \le M_n^{r,k}(t) \le H_n^k(t)$. Also it is shown in both cases that $M_n^{r,k}(t)$ is a decreasing function of r. We show that $H_{n-1}^{k-1}(t) \le H_n^k(t) \le H_{n-1}^k(t)$, if T_n is independent of $T_1, \ldots, T_{n-1}$.

Using the concept of signature introduced by Samaniego (1985), the MRL function of a coherent structure with exchangeable components is considered in Section 3. A brief conclusion remarks is given in Section 4.

2. MRL function of a $(n-k+1)$-out-of-n system with nonidentical or exchangeable components

Let $T_1, T_2, \ldots, T_n$ denote the lifetimes of n components which are connected in a $(n-k+1)$-out-of-n structure, $k = 1, 2, \ldots, n$. Suppose furthermore, that T_i's are nonnegative and absolutely continuous random variables. in IID case when T_i's are independent and have a common distribution function F and survival function $\bar{F} = 1 - F$, Asadi and Goliforushani (2008) showed that $M_n^{r,k}(t)$ can be written as a convex combination of $H_{n-i}^{k-i}(t)$ as follow

$$M_n^{r,k}(t) = E(T_{k:n} - t | T_{r:n} > t) = \sum_{i=0}^{r-1} P_i(t) H_{n-i}^{k-i}(t) \tag{1}$$

where $P_i(t) = \frac{\binom{n}{i}\phi^i(t)}{\sum_{j=0}^{r-1}\binom{n}{j}\phi^i(t)}$, $\phi(t) = \frac{F(t)}{\bar{F}(t)}$ and $H_{n-i}^{k-i}(t) = E(T_{k-i:n-i} - t | T_{1:n-i} > t)$.

When T_i's are independent but not identical, the similar expression for $M_n^{r,k}(t)$ is obtained in Theorem 1. Suppose $F_i(x) = 1 - \bar{F}_i(x)$ is the distribution function of T_i. We assume that $\cap_{i=1}^n (a_i, b_i) \neq \emptyset$ where a_i and b_i are left extremity and right extremity of F_i, respectively. In other words, we assume that $a < b$ where $a = max\{a_i, i = 1, \ldots, n\}$ and $b = min\{b_i, i = 1, \ldots, n\}$ and then define $M_n^{r,k}(t)$ for any $a < t < b$.

Theorem 1. In nonidentical case, for $a < t < b$ we have

$$M_n^{r,k}(t) = E(T_{k:n} - t | T_{r:n} > t) = \sum_{i=0}^{r-1} \sum_{C_i} P_{C_i}(t) H_{C_i'}^{k-i}(t) \tag{2}$$

where $P_{C_i}(t) = \frac{\phi_{C_i}(t)}{\sum_{j=0}^{r-1}\sum_{C_j}\phi_{C_j}(t)}$, $\phi_{C_j}(t) = \prod_{l \in C_j} \frac{F_l(t)}{\bar{F}_l(t)}$, C_j is a subset of $C = \{1, 2, \ldots, n\}$ with cardinality j, $C_j' = C - C_j$ and $H_{C_i'}^{k-i}(t) = E(T_{k-i:n-i} - t | T_{1:n-i} >$

t) is the MRL function of a $(n-k+1)$-out-of-$(n-i)$ subsystem consisting of components belong to C_i'. We assume that $\phi_\emptyset(t)=1$ where $\emptyset$ is the empty set.

Proof. Using the joint distribution of $T_{k:n}$ and $T_{r:n}$ one can show that

$$P(T_{k:n}-t>x|T_{r:n}>t)=\frac{P(T_{k:n}>t+x,T_{r:n}>t)}{P(T_{r:n}>t)}$$

$$=\frac{\sum_{i=0}^{r-1}\sum_{C_i}\prod_{l\in C_i}F_l(t)\sum_{u=0}^{k-i-1}\sum_{C'_{i(u)}}\prod_{l\in C'_{i(u)}}[\bar{F}_l(t)-\bar{F}_l(t+x)]\prod_{l\in C_i'-C'_{i(u)}}\bar{F}_l(t+x)}{\sum_{j=0}^{r-1}\sum_{C_j}\prod_{l\in C_j}F_l(t)\prod_{l\in C_j'}\bar{F}_l(t)}$$

where $C'_{i(u)}$ is a subset of $C_i'=C-C_i$ with cardinality u. Hence we have

$$M_n^{r,k}(t)=\int_0^\infty P(T_{k:n}-t>x|T_{r:n}>t)dx=\int_t^\infty P(T_{k:n}>x|T_{r:n}>t)dx$$

$$=\int_t^\infty\frac{\sum_{i=0}^{r-1}\sum_{C_i}\prod_{l\in C_i}F_l(t)\prod_{l\in C_i'}\bar{F}_l(t)\sum_{u=0}^{k-i-1}\sum_{C'_{i(u)}}\prod_{l\in C'_{i(u)}}[1-\theta_{t,l}(x)]\prod_{l\in C_i'-C'_{i(u)}}\theta_{t,l}(x)}{\sum_{j=0}^{r-1}\sum_{C_j}\prod_{l\in C_j}F_l(t)\prod_{l\in C_j'}\bar{F}_l(t)}dx$$

where $\theta_{t,l}(x)=\frac{\bar{F}_l(x)}{\bar{F}_l(t)}$, $x>t$ and we assume that multiplication over the empty set is 1. Using the expression of $P_{C_i}(t)$ we can write

$$M_n^{r,k}(t)=\sum_{i=0}^{r-1}\sum_{C_i}P_{C_i}(t)\int_t^\infty\sum_{u=0}^{k-i-1}\sum_{C'_{i(u)}}\prod_{l\in C'_{i(u)}}[1-\theta_{t,l}(x)]\prod_{l\in C_i'-C'_{i(u)}}\theta_{t,l}(x)dx \quad (3)$$

If we take $r=1$ in equation (3) we get
$H_n^k(t)=M_n^{1,k}(t)=E(T_{k:n}-t|T_{1:n}>t)=\int_t^\infty\sum_{u=0}^{k-1}\sum_{C_u}\prod_{l\in C_u}[1-\theta_{t,l}(x)]\prod_{l\in C_u'}\theta_{t,l}(x)dx$
and hence if we consider the components belong to C_i' and we replace in above equation n and k by $n-i$ and $k-i$, respectively, we then have
$H_{C_i'}^{k-i}(t)=E(T_{k-i:n-i}-t|T_{1:n-i}>t)=\int_t^\infty\sum_{u=0}^{k-i-1}\sum_{C'_{i(u)}}\prod_{l\in C'_{i(u)}}[1-\theta_{t,l}(x)]\prod_{l\in C_i'-C'_{i(u)}}\theta_{t,l}(x)dx$.
Using this the equation (3) is simply reduces to the equation (2) and the proof of the theorem is completed.

We now consider the case where T_i's are exchangeable random variables that is all $n!$ permutations of $T_1,\ldots,T_n$ have the same joint distribution. Hence T_i's are

identically distributed and have a common marginal distribution. The following theorem gives some properties of $H_n^k(t) = E(T_{k:n} - t|T_{1:n} > t)$ when T_i's are exchangeable random variables.

Theorem 2. Suppose $T_1, \ldots, T_n$ are exchangeable random variables. We have
(a) $E(T_1 - t|T_{1:n} > t) = 1/n \sum_{k=1}^n H_n^k(t)$, $n = 1, 2, \ldots$
(b) The following recurrence relation holds for $H_n^k(t)$.

$$nH_{n-1,n}^k(t) = (n-k)H_n^k(t) + kH_n^{k+1}(t), k = 1, 2, \ldots, n-1$$

where $H_{n-1,n}^k(t) = E(T_{k:n-1} - t|T_{1:n} > t)$.

Proof. (a) Let $T_t^{1,k,n} = T_{k:n} - t|T_{1:n} > t$ denote the residual lifetime of a $(n-k+1)$-out-of-n system under the condition that all components of the system are working at time t. We have

$$P(T_t^{1,k,n} > x) = \frac{P(T_{k:n} > t + x, T_{1:n} > t)}{P(T_{1:n} > t)}$$

$$= \frac{\sum_{i=0}^{k-1} \binom{n}{i} P(t < T_1 < t + x, \ldots, t < T_i < t + x, T_{i+1} > t + x, \ldots, T_n > t + x)}{P(T_1 > t, T_2 > t, \ldots, T_n > t)}. \quad (4)$$

This shows that when T_i's are exchangeable random variables, $T_t^{1,k,n} \overset{\text{st}}{=} T_t^{k:n}$ where $T_t^{k:n}$ is the kth order statistics from the sample consisting of $T_t^i = T_i - t|T_{1:n} > t$, $i = 1, 2, \ldots, n$ which are exchangeable random variables and $\overset{\text{st}}{=}$ stands for distribution. We have
$H_n^k(t) = ET_t^{1,k,n} = ET_t^{k:n}$ and we get $\sum_{k=1}^n H_n^k(t) = \sum_{k=1}^n ET_t^{k:n}$. But we note that $\sum_{k=1}^n T_t^{k:n} = \sum_{k=1}^n T_t^k = \sum_{k=1}^n (T_k - t|T_{1:n} > t)$, a.s., therefore $\sum_{k=1}^n ET_t^{k:n} = \sum_{k=1}^n E(T_k - t|T_{1:n} > t)$, that is $\sum_{k=1}^n H_n^k(t) = nE(T_1 - t|T_{1:n} > t)$ as T_t^k, $k = 1, 2, \ldots, n$ are identically distributed. This completes the proof of part (a).
(b) We note that if T_i's are exchangeable then $P(T_{\pi_{(1)}} < T_{\pi_{(2)}} < \cdots < T_{\pi_{(n)}}) = 1/n!$, for any permutation π of the numbers $\{1, 2, \ldots, n\}$. Hence we have $P(T_{r:n} = T_i) = 1/n$, $i, r = 1, 2, \ldots, n$. In view of this the equation (3.4.1) in David and Nagaraja (2003, page 44) is also true for exchangeable random variables. If we use this equation for the exchangeable random variables $T_t^i = T_i - t|T_{1:n} > t$, $i = 1, 2, \ldots, n$, we have
$nP(T_{k:n-1} - t < x|T_{1:n} > t) = kP(T_{k+1:n} - t < x|T_{1:n} > t) + (n-k)P(T_{k:n} - t < x|T_{1:n} > t)$.
It simply implies that
$nP(T_{k:n-1} - t > x|T_{1:n} > t) = kP(T_{k+1:n} - t > x|T_{1:n} > t) + (n-k)P(T_{k:n} - t > x|T_{1:n} > t)$.
By integrating both sides of the above equation with respect to x, the proof of the theorem follows.

Remark 1. In IID case when T_i's are independent and have a common distribution function F, we note that $E(T_1 - t|T_{1:n} > t) = E(T_1 - t|T_1 > t) = m_F(t)$. Hence $m_F(t) = 1/n \sum_{k=1}^{n} H_n^k(t)$, a result obtained by Asadi and Goliforushani (2008). In this case we also note that $H_{n-1,n}^k(t) = E(T_{k:n-1} - t|T_{1:n} > t) = E(T_{k:n-1} - t|T_{1:n-1} > t) = H_{n-1}^k(t)$. It seems that the argument given in the proof of part (a) of Theorem 2 is simpler than that of given in Asadi and Goliforushani (2008) for IID case, as the integrating operations are not required.

The following lemma gives a result on $H_n^k(t)$ in nonidentical case which extends the corresponding result in IID case, $m_F(t) = 1/n \sum_{k=1}^{n} H_n^k(t)$, obtained by Asadi and Goliforushani (2008).

Lemma 1. In nonidentical case we have

$$\sum_{k=1}^{n} m_{F_k}(t) = \sum_{k=1}^{n} H_n^k(t)$$

where $m_{F_k} = E(T_k - t|T_k > t)$ and F_k is the distribution function of T_k, $k = 1, \ldots, n$.

Proof. Since T_i's are independent we have

$$P(T_t^{1,k,n} > x) = \frac{P(T_{k:n} > t + x, T_{1:n} > t)}{P(T_{1:n} > t)}$$

$$= \frac{\sum_{i=0}^{k-1} \sum_{C_i} \prod_{l \in C_i} [\bar{F}_l(t) - \bar{F}_l(t+x)] \prod_{l \in C_i'} \bar{F}_l(t+x)}{\prod_{i=1}^{n} \bar{F}_i(t)} = P(T_t^{k:n} > x) \quad (5)$$

where $T_t^{1,k,n} = T_{k:n} - t|T_{1:n} > t$ and $T_t^{k:n}$ here is the kth order statistics from the sample consisting of $T_i - t|T_i > t$, $i = 1, \ldots, n$ when T_i's are independent but not identical. We note that $T_t^{1,k,n} \stackrel{\text{st}}{=} T_t^{k:n}$ hence we have $E(T_t^{1,k,n}) = E(T_t^{k:n})$ and $\sum_{k=1}^{n} T_t^{k:n} = \sum_{k=1}^{n} (T_k - t|T_k > t)$, a.s., therefore $\sum_{k=1}^{n} E(T_t^{1,k,n}) = \sum_{k=1}^{n} E(T_t^{k:n}) = E \sum_{k=1}^{n} (T_k - t|T_k > t)$, that is $\sum_{k=1}^{n} H_n^k(t) = \sum_{k=1}^{n} m_{F_k}(t)$. This competes the proof of the lemma.

We now give some other results on $H_n^k(t)$ in both nonidentical and exchangeable cases.

Theorem 3. $H_n^k(t) = E(T_{k:n} - t|T_{1:n} > t)$ is a decreasing function of n in nonidentical case and in exchangeable case we have $H_n^k(t) \leq H_{n-1,n}^k(t) = E(T_{k:n-1} - t|T_{1:n} > t)$.

Proof. From Equation (5), in nonidentical case we can write

$$P(T_t^{1,k,n} > x) = \sum_{i=0}^{k-1} \sum_{C_i} \prod_{l \in C_i} \frac{\bar{F}_l(t) - \bar{F}_l(t+x)}{\bar{F}_l(t)} \prod_{l \in C_i'} \frac{\bar{F}_l(t+x)}{\bar{F}_l(t)} = P(\sum_{j=1}^{n} Z_{t,x}^j \leq k-1) \quad (6)$$

where $Z_{t,x}^j$, $j = 1, \ldots, n$ are independent random variables and $Z_{t,x}^j$ is distributed as Binomial $(1, \frac{\bar{F}_j(t)-\bar{F}_j(t+x)}{\bar{F}_j(t)})$. It is easy to see that

$$P(T_t^{1,k,n} > x) - P(T_t^{1,k,n-1} > x) = P(\sum_{j=1}^{n} Z_{t,x}^j \leq k-1) - P(\sum_{j=1}^{n-1} Z_{t,x}^j \leq k-1)$$

$$= P(Z_{t,x}^n = 1)[P(\sum_{j=1}^{n-1} Z_{t,x}^j \leq k-2) - P(\sum_{j=1}^{n-1} Z_{t,x}^j \leq k-1)] \leq 0.$$

Hence we have

$$H_n^k(t) - H_{n-1}^k(t) = \int_0^\infty [P(T_t^{1,k,n} > x) - P(T_t^{1,k,n-1} > x)]dx \leq 0.$$

In exchangeable case from Part (b) of Theorem 2, we have $nH_{n-1,n}^k(t) = (n-k)H_n^k(t)+kH_n^{k+1}(t)$. It implies that $H_n^k(t)-H_{n-1,n}^k(t) = k/n[H_n^k(t)-H_n^{k+1}(t)] \leq 0$, since $H_n^k(t)-H_n^{k+1}(t) = E(T_{k:n}-T_{k+1:n}|T_{1:n} > t)$ which is obviously non positive. This implies that $H_n^k(t) \leq H_{n-1,n}^k(t) = E(T_{k:n-1} - t|T_{1:n} > t)$.

Remark 2. Let $(T_1, \ldots, T_{n-1})$ be an exchangeable random vector and T_n is independent of $(T_1, T_2, \ldots, T_{n-1})$, it then can be easily shown that $T_{k:n-1} - t|T_{1:n} > t \stackrel{\text{st}}{=} T_{k:n-1} - t|T_{1:n-1} > t$ and hence $H_{n-1,n}^k(t) = E(T_{k:n-1} - t|T_{1:n} > t) = E(T_{k:n-1} - t|T_{1:n-1} > t) = H_{n-1}^k(t)$. Therefore in view of Theorem 3 we have $H_n^k(t) \leq H_{n-1}^k(t)$.

The following lemma shows in both described cases, that $H_n^k(t)$ is expressible in terms of simpler MRL functions of the minimum in samples of sizes $n-k+1, \ldots, n$.

Lemma 2. (a) In exchangeable case we have

$$H_n^k(t) = ET_t^{1,k,n} = E(T_{k:n} - t|T_{1:n} > t)$$
$$= \sum_{j=n-k+1}^{n} (-1)^{j-n+k-1} \binom{n}{j} \binom{j-1}{n-k} H_{j,n}^1(t)$$

where

$$H_{j,n}^1(t) = E(T_{1:j} - t|T_{1:n} > t) = \int_0^\infty \frac{\bar{F}(\overbrace{t+x, \ldots, t+x}^{j}, \overbrace{t, \ldots, t}^{n-j})}{\bar{F}(t, \ldots, t)} dx,$$

and $\bar{F}(x_1, \ldots, x_n) = P(T_1 > x_1, \ldots, T_n > x_n)$.

(b) In nonidentical case we have

$$H_n^k(t) = E(T_{k:n} - t|T_{1:n} > t) = \sum_{j=n-k+1}^{n} (-1)^{j-n+k-1} \binom{j-1}{n-k} \sum_{C_j} H_{C_j}^1(t)$$

where

$$H^1_{C_j}(t) = E(minT_l - t | minT_l > t, l \in C_j) = \int_0^\infty \prod_{l \in C_j} \frac{\bar{F}_l(t+x)}{\bar{F}_l(t)} dx$$

is the MRL of a series system consisting of components belongs to C_j and C_j is a subset of $C = \{1, \ldots, n\}$ with cardinality j.

Proof. (a) Using Equation (3.4.3) in David and Nagaraja (2003, page 46) we have

$$P(T_{k:n} - t > x | T_{1:n} > t) = \sum_{j=n-k+1}^{n} (-1)^{j-n+k-1} \binom{n}{j} \binom{j-1}{n-k} P(T_{1:j} - t > x | T_{1:n} > t).$$

By integrating both sides of the above equation with respect to x, the proof of part (a) follows.

(b) Using Equation (3.4.2) in David and Nagaraja (2003, page 46), we can write

$$\begin{aligned} &P(T_{k:n} - t > x | T_{1:n} > t) \\ &= \sum_{j=n-k+1}^{n} (-1)^{j-n+k-1} \binom{j-1}{n-k} \sum_{C_j} \frac{\prod_{l \in C_j} \bar{F}_l(t+x) \prod_{l \in C'_j} \bar{F}_l(t)}{\prod_{i=1}^{n} \bar{F}_i(t)} \\ &= \sum_{j=n-k+1}^{n} (-1)^{j-n+k-1} \binom{j-1}{n-k} \sum_{C_j} P(minT_l > t + x | minT_l > t, l \in C_j). \end{aligned}$$

Again by integrating both sides with respect to x the proof follows.

Remark 3. In view of Equation (2) and part (b) of Lemma 2, we note in non-identical case that:

$$\begin{aligned} M^{r,k}_n(t) &= E(T_{k:n} - t | T_{r:n} > t) \\ &= \sum_{i=0}^{r-1} \sum_{C_i} P_{C_i}(t) \sum_{j=n-k+1}^{n} (-1)^{j-n+k-1} \binom{j-1}{n-k} \sum_{C'_{i(j)}} H^1_{C'_{i(j)}}(t) \quad (7) \end{aligned}$$

where $C'_{i(j)}$ is a subset of $C'_i = C - C_i$ with cardinality j and $H^1_{C'_{i(j)}}(t)$ is similarly defined to $H^1_{C_i}(t)$ which is defined in Lemma 2. The Equation (7) shows in non-identical case that $M^{r,k}_n(t)$, as well as $H^k_n(t)$, can also be expressed in terms of simpler MRL of the minimum in samples of sizes $n-k+1, \ldots, n$. Using Part (a) of Lemma 2 we can simply obtain an expression for $H^k_n(t) = E(T_{k:n} - t | T_{1:n} > t)$ in terms of the joint survival function of T_i's, $\bar{F}(x_1, \ldots, x_n) = P(T_1 > x_1, \ldots, T_n > x_n)$ which is useful in practical applications.

Remark 4. From Equation (4), in exchangeable case we can write

$$H^1_n(t) = E(T_{1:n} - t | T_{1:n} > t) = \int_0^\infty \frac{\bar{F}(t+x, \ldots, t+x)}{\bar{F}(t, \ldots, t)} dx,$$

which is easily decreasing (increasing) in t, if $\frac{\bar{F}(t+x,\ldots,t+x)}{\bar{F}(t,\ldots,t)}$ is decreasing (increasing) in t for $x \geq 0$. A sufficient condition for this is that the joint distribution of T_i's be a multivariate IFR (DFR) distribution (see, for example, Barlow and Proschan, (1975)).

Similarly in nonidentical case we have $H_n^1(t) = \int_0^\infty \prod_{j=1}^n \frac{\bar{F}_j(t+x)}{\bar{F}_j(t)} dx$ which is obviously decreasing (increasing) in t, if F_j's $j = 1, \ldots, n$ are IFR (DFR) distributions. We recall that a distribution F_j is said to be an IFR (DFR) distribution function if the corresponding failure rate, $r_{F_j}(t) = \frac{f_j(t)}{\bar{F}_j(t)}$ is increasing (decreasing) in t. In this case and in view of Equation (6) we have

$$H_n^k(t) = E(T_{k:n} - t | T_{1:n} > t) = \int_0^\infty P(\sum_{j=1}^n Z_{t,x}^j \leq k-1) dx$$

where $Z_{t,x}^j$, $j = 1, \ldots, n$ are independent random variables and $Z_{t,x}^j$ is distributed as Binomial(1,1-$\frac{\bar{F}_j(t+x)}{\bar{F}_j(t)}$). It is easy to show that $H_n^k(t)$ is decreasing (increasing) in t, if 1-$\frac{\bar{F}_j(t+x)}{\bar{F}_j(t)}$, $j = 1, 2, \ldots, n$ are increasing (decreasing) in t or equivalently F_j's are IFR (DFR) distributions.

In the following lemma we obtain an upper bound and a lower bound for $M_n^{r,k}(t)$ in nonidentical case.

Lemma 3. In nonidentical case we have

$$min_{C'_{r-1}} H_{C'_{r-1}}^{k-r+1}(t) \leq M_n^{r,k}(t) \leq H_n^k(t).$$

Proof. We first show that $H_n^k(t) - H_{n-1}^{k-1}(t) \geq 0$. We have

$$P(T_{k:n} - t > x | T_{1:n} > t) = P(\sum_{j=1}^n Z_{t,x}^j \leq k-1)$$

where $Z_{t,x}^j$ is defined in Remark 4. It is easy to show that

$$\begin{aligned}
&H_n^k(t) - H_{n-1}^{k-1}(t) \\
&\quad = \int_0^\infty \left(P(T_{k:n} - t > x | T_{1:n} > t) - P(T_{k-1:n-1} - t > x | T_{1:n-1} > t) \right) dx \\
&\quad = \int_0^\infty \left(P(\sum_{j=1}^n Z_{t,x}^j \leq k-1) - P(\sum_{j=1}^{n-1} Z_{t,x}^j \leq k-2) \right) dx \\
&\quad = \int_0^\infty \frac{\bar{F}_n(t+x)}{\bar{F}_n(t)} P(\sum_{j=1}^{n-1} Z_{t,x}^j = k-1) dx
\end{aligned}$$

which is obviously nonnegative. Using this and from Equation (2) we have

$$M_n^{r,k}(t) = E(T_{k:n} - t|T_{r:n} > t)$$
$$= \sum_{i=0}^{r-1} \sum_{C_i} P_{C_i}(t) H_{C_i'}^{k-i}(t) \leq \sum_{i=0}^{r-1} \sum_{C_i} P_{C_i}(t) H_n^k(t) = H_n^k(t).$$

For each $0 \leq i \leq r-1$ and C_i there exists C_{r-1} such that $C_{r-1}' \subseteq C_i'$. As $H_n^k(t) \geq H_{n-1}^{k-1}(t)$ we have $H_{C_i'}^{k-i}(t) \geq H_{C_{r-1}'}^{k-r+1}(t)$. Hence we can write

$$M_n^{r,k}(t) = \sum_{i=0}^{r-1} \sum_{C_i} P_{C_i}(t) H_{C_i'}^{k-i}(t)$$
$$\geq \sum_{i=0}^{r-1} \sum_{C_i} P_{C_i}(t) min_{C_{r-1}'} H_{C_{r-1}'}^{k-r+1}(t) = min_{C_{r-1}'} H_{C_{r-1}'}^{k-r+1}(t).$$

This completes the proof of the lemma.

In IID case we note that

$$min_{C_{r-1}'} H_{C_{r-1}'}^{k-r+1}(t) = H_{n-r+1}^{k-r+1}(t) = E(T_{k-r+1:n-r+1} - t|T_{1:n-r+1} > t).$$

Lemma 4. If T_i, $i = 1, 2, \ldots, n-1$ are exchangeable random variables and T_n is independent of $(T_1, T_2, \ldots, T_{n-1})$ then $H_n^k(t) \geq H_{n-1}^{k-1}(t)$.

Proof. In view of the proof of Part (a) of Theorem 2, we note that

$$T_{k:n} - t|T_{1:n} > t \stackrel{\text{st}}{=} T_t^{k:n}$$

where $T_t^{k:n}$ is the kth order statistics from the sample consisting of $T_t^{i,n} = T_i - t|T_{1:n} > t$, $i = 1, 2, \ldots, n$ in which $T_t^{i,n}$ for $i = 1, 2, \ldots, n-1$ are exchangeable random variables with common survival function

$$\bar{G}_{n,t}(x) = P(T_t^{i,n} > x) = \frac{P(T_1 > t + x, T_2 > t, \ldots, T_n > t)}{P(T_1 > t, T_2 > t, \ldots, T_n > t)}.$$

$T_t^{n,n}$ is independent of $(T_t^{1,n}, T_t^{2,n}, \ldots, T_t^{n-1,n})$ and its survival function is $\frac{P(T_n > t+x)}{P(T_n > t)}$. Since T_n is independent of $(T_1, T_2, \ldots, T_{n-1})$ we have $\bar{G}_{n,t}(x) = \bar{G}_{n-1,t}(x)$. In other words $T_t^{i,n} \stackrel{\text{st}}{=} T_t^{i,n-1}$, $i = 1, 2, \ldots, n-1$. We also note that

$$P(T_{k:n} - t > x|T_{1:n} > t) = P(T_t^{k:n} > x) = P(\sum_{j=1}^{n} W_{t,x}^{j,n} \leq k-1)$$

where $W_{t,x}^{j,n}$, $j = 1, 2, \ldots, n-1$ are exchangeable random variables and $W_{t,x}^{j,n}$ is distributed as Binomial(1,1- $\bar{G}_{n,t}(x)$). $W_{t,x}^{n,n}$ is independent of $W_{t,x}^{j,n}$, $j = 1, 2, \ldots, n-1$

and is distributed as Binomial(1,1-$\frac{P(T_n>t+x)}{P(T_n>t)}$). Note that $W_{t,x}^{j,n} \overset{\text{st}}{=} W_{t,x}^{j,n-1}$, $j = 1, 2, \ldots, n-1$. Now from above discussion it is easy to show that

$$\begin{aligned} P(T_{k:n} - t > x | T_{1:n} > t) - P(T_{k-1:n-1} - t > x | T_{1:n-1} > t) \\ &= P(T_t^{k:n} > x) - P(T_t^{k-1:n-1} > x) \\ &= P(\sum_{j=1}^{n} W_{t,x}^{j,n} \leq k-1) - P(\sum_{j=1}^{n-1} W_{t,x}^{j,n-1} \leq k-2) \\ &= P(W_{t,x}^{n,n} = 0) P(\sum_{j=1}^{n-1} W_{t,x}^{j,n-1} = k-1) \end{aligned}$$

which is obviously nonnegative. By integrating both sides with respect to x the proof of the lemma follows.

Remark 5. In view of Remark 2 and Lemma 4, under the condition that T_n is independent of exchangeable random vector $(T_1, T_2, \ldots, T_{n-1})$ we have

$$H_{n-1}^{k-1}(t) \leq H_n^k(t) \leq H_{n-1}^k(t).$$

Although the above inequality is always true in nonidentical case, however it is not generally hold in exchangeable case. See the following example.

Example. Suppose that the joint distribution of $T_1, \ldots, T_n$ is Marshal and Olkin's multivariate exponential with survival function

$$\begin{aligned} &\bar{F}(x_1, \ldots, x_n) \\ &= exp\{-\sum_{i=1}^{n} \lambda_i x_i - \sum \sum_{i_1 < i_2} \lambda_{i_1, i_2} max(x_{i_1}, x_{i_2}) - \cdots - \lambda_{12\ldots n} max(x_1, \ldots, x_n)\}. \end{aligned}$$

For the special case $\lambda_1 = \ldots = \lambda_n = \lambda_{12} = \ldots = \lambda_{12\ldots n} = \lambda$, $\bar{F}(x_1, \ldots, x_n)$ is exchangeable. It can be simply shown that

$$\frac{\bar{F}(\overbrace{t+x, \ldots, t+x}^{j}, \overbrace{t, \ldots, t}^{n-j})}{\bar{F}(t, \ldots, t)} = exp\{-(2^n - 2^{n-j})\lambda x\}$$

and hence using Part (a) of Lemma 2 we have

$$\begin{aligned} H_n^k(t) &= \lambda^{-1} \sum_{i=n-k+1}^{n} \frac{(-1)^{i+k-n-1}}{2^n - 2^{n-i}} \binom{n}{i} \binom{i-1}{n-k} \\ &= \lambda^{-1} \sum_{i=0}^{k-1} \frac{(-1)^{k-i-1}}{2^n - 2^i} \binom{n}{i} \binom{n-i-1}{n-k}, \end{aligned}$$

which is a positive constant. For $\lambda = 1$, $n = 3$ and $k = 2$ we have $H_n^k(t) = 0.21$, $H_{n-1}^k(t) = 0.66$ and $H_{n-1}^{k-1}(t) = 0.33$. It shows that the inequality given in Remark 5 is not hold and the assumption T_n is independent of $(T_1, \ldots, T_{n-1})$ is required. Now in nonidentical case we will show in the following theorem, that $M_n^{r,k}(t) = E(T_{k:n} - t | T_{r:n} > t)$, $1 \le r \le k \le n$ is decreasing in t, if F_j's are IFR distributions.

Theorem 4. In nonidentical case if F_i, $i = 1, \ldots, n$ are IFR distributions then $M_n^{r,k}(t) = E(T_{k:n} - t | T_{r:n} > t)$, $1 \le r \le k \le n$ is a decreasing function of t.

Proof. From Equation (2) we have
$M_n^{r,k}(t) = E(T_{k:n} - t | T_{r:n} > t) = \sum_{i=0}^{r-1} \sum_{C_i} P_{C_i}(t) H_{C_i'}^{k-i}(t)$ where $P_{C_i}(t) = \frac{\phi_{C_i}(t)}{\sum_{j=0}^{r-1} \sum_{C_j} \phi_{C_j}(t)}$ and $\phi_{C_j}(t) = \prod_{l \in C_j} \frac{F_l(t)}{F(t)}$. Since $\phi_{C_j}(t)$ is an increasing function of t, without loss of generality we assume that $\phi_{C_j}(t) = t^j$. Hence we show that

$$M(t) = \frac{\sum_{i=0}^{r-1} \sum_{C_i} t^i H_{C_i'}^{k-i}(t)}{\sum_{j=0}^{r-1} \binom{n}{j} t^j}$$

is a decreasing function of t. We have, on differentiating both sides of this last equation,

$$M'(t) = \frac{1}{\left(\sum_{j=0}^{r-1} \binom{n}{j} t^j\right)^2} \left(\sum_{i=0}^{r-1} \sum_{C_i} i t^{i-1} H_{C_i'}^{k-i}(t) \sum_{j=0}^{r-1} \binom{n}{j} t^j \right.$$
$$\left. - \sum_{j=0}^{r-1} j \binom{n}{j} t^{j-1} \sum_{i=0}^{r-1} \sum_{C_i} t^i H_{C_i'}^{k-i}(t) + \sum_{i=0}^{r-1} \sum_{C_i} t^i (H_{C_i'}^{k-i}(t))' \sum_{j=0}^{r-1} \binom{n}{j} t^j \right). \quad (8)$$

As stated in Remark 4, $H_{C_i'}^{k-i}(t)$ is decreasing in t, since F_j's are IFR distributions, and hence the last term on the right hand side of the Equation (8) is non positive. We show that the difference between the first and the second terms is also non positive. Let we denote this difference by D. After some algebraic manipulations it can be shown that

$$D = \frac{1}{\left(\sum_{j=0}^{r-1} \binom{n}{j} t^j\right)^2} \sum_{i=0}^{r-1} \sum_{C_i} \sum_{j=i}^{r-1} (j - i) \sum_{C_j} \left(H_{C_j'}^{k-j}(t) - H_{C_i'}^{k-i}(t) \right) t^{i+j-1}.$$

We know that for each C_j $(j = 1, \ldots, n)$ there exist subsets $C_1, C_2, \ldots, C_{j-1}$ such that $C_j' \subset C_i'$, $i = 1, 2, \ldots, j-1$. It is shown in the proof of Lemma 3, that $H_n^k(t) - H_{n-1}^{k-1}(t) \ge 0$, hence D is non positive and this completes the proof of the theorem.

Remark 6. In view of the proof of Theorem 4 we can conclude in nonidentical case, that if the MRLs $H_{C_i'}^{k-i}(t)$, $i = 1, 2, \ldots, r-1$ are decreasing functions of time then the MRL $M_n^{r,k}(t)$ is also decreasing in time. We have already seen in Remark

4, that if F_i, $i = 1, \ldots, n$ are DFR distributions then $H_n^k(t)$ is increasing in t. This result, however, is not generally true for $M_n^{r,k}(t)$ (see Asadi and Goliforushani (2008) for a counterexample in IID case).

Khanjari(2008) showed in nonidentical case that $M_n^{r,n}(t) = E(T_{n:n} - t|T_{r:n} > t)$, $1 \leq r \leq n$ is decreasing in r. Now using a different argument and in both nonidentical and exchangeable cases we show in the following theorem that $M_n^{r,k}(t) = E(T_{k:n} - t|T_{r:n} > t)$, $1 \leq r \leq k \leq n$ is also a decreasing function of r.

Theorem 5. In both described cases we have $M_n^{r-1,k}(t) \geq M_n^{r,k}(t)$, $2 \leq r \leq k \leq n$.

Proof. We note that $T_{k:n} - t|T_{r:n} > t \stackrel{\text{st}}{=} Z_{k:n}^{t,r}$ where $Z_{k:n}^{t,r}$ is the kth order statistics from the sample consisting of $Z_{i,t,r} = T_i - t|T_{r:n} > t$, $i = 1, \ldots, n$. We first show that $Z_{i,t,r} \leq_{st} Z_{i,t,r-1}$. Suppose $x < 0$ is given and $Z_{i,t,r-1} \leq x$, that is $T_i - t \leq x|T_{r-1:n} > t$. This means T_i is ordered between the first $r-2$ smallest of $T_1, \ldots, T_n$. Obviously T_i is also between the first $r-1$ smallest of T_j's. Hence if $T_{r:n} > t$ then $T_i \leq t + x$ that is $Z_{i,t,r} \leq x$. Now suppose $x > 0$ and $Z_{i,t,r} > x$ that is $T_i - t > x|T_{r:n} > t$. It means T_i is between the last $n-r+1$ and consequently $n-r+2$ largest values of $T_1, \ldots, T_n$. This implies that if $T_{r-1:n} > t$ then $T_i > t + x$ that is $Z_{i,t,r-1} > x$. Therefore $Z_{i,t,r} \leq_{st} Z_{i,t,r-1}$. We now show that $Z_{k:n}^{t,r} \leq_{st} Z_{k:n}^{t,r-1}$. Suppose $x > 0$ and $Z_{k:n}^{t,r} > x$. That is at most $k-1$ out of $Z_{i,t,r}$, $1 \leq i \leq n$ are less than x. We show that $Z_{k:n}^{t,r-1} > x$. Suppose not that is at least k out of $Z_{i,t,r-1}$, $1 \leq i \leq n$ are less than x. Since $Z_{i,t,r} \leq_{st} Z_{i,t,r-1}$, $1 \leq i \leq n$, there exist at least k out of $Z_{i,t,r}$, $1 \leq i \leq n$ which are less than x and this is a contradiction. Therefore $Z_{k:n}^{t,r-1} > x$. Similarly for $x < 0$ it can be shown that the event $Z_{k:n}^{t,r-1} < x$ is a subset of the event $Z_{k:n}^{t,r} < x$. Hence $Z_{k:n}^{t,r} \leq_{st} Z_{k:n}^{t,r-1}$ and we get $M_n^{r,k}(t) = EZ_{k:n}^{t,r} \leq EZ_{k:n}^{t,r-1} = M_n^{r-1,k}(t)$. This completes the proof of the theorem.

Remark 7. In both nonidentical and exchangeable cases and in view of Theorem 5, for $t > 0$ and $1 \leq r \leq k \leq n$ we have

$$\begin{aligned} M_n^{k,k}(t) = E(T_{k:n} - t|T_{k:n} > t) \leq M_n^{r,k}(t) = E(T_{k:n} - t|T_{r:n} > t) \leq M_n^{1,k}(t) \\ = E(T_{k:n} - t|T_{1:n} > t) \end{aligned}$$

where $M_n^{1,k}(t) = H_n^k(t)$.

3. Mean residual life function of a coherent system

In this section we consider a coherent system consisting of n components. Let T_i, $i = 1, 2, \ldots, n$ and $T = \phi(T_1, \ldots, T_n)$ which are nonnegative and absolutely continuous random variables, denote the lifetimes of components and the system, respectively. Samaniego (1985) introduced the concept of the signature of ϕ as the probability vector $\mathbf{p} = (p_1, \ldots, p_n)$ where $p_i = P\{\phi(T_1, \ldots, T_n) = T_{i:n}\}$ and $T_{i:n}$, $i = 1, 2, \ldots, n$ represent ordered lifetimes of the components. In IID case Kochar

et al.(1999) showed that

$$P(T > t) = \sum_{i=1}^{n} p_i P(T_{i:n} > t). \tag{9}$$

Navarro et al.(2005) have shown that this relation is also true in exchangeable case. It should be noted in nonidentical case that the Equation (9) is not generally true (see a counterexample in Navarro et al.(2008)). In both IID and exchangeable cases, one can show that $p_i = \frac{n_i}{n!}$ in which n_i is the number of permutations of $T_1, \ldots, T_n$ such that $T = \phi(T_1, \ldots, T_n) = T_{i:n}$.

Here we consider the mean residual life of the system with exchangeable components, when all components of the system are working at time t, $MR_n^1(t) = E(T - t|T_{1:n} > t)$. From Equation (9), $MR_n^1(t)$ can be written as follow

$$\begin{aligned} MR_n^1(t) &= \int_0^\infty P(T - t > x|T_{1:n} > t)dx = \sum_{i=1}^{n} p_i \int_0^\infty P(T_{i:n} - t > x|T_{1:n} > t)dx \\ &= \sum_{i=1}^{n} p_i H_n^i(t). \end{aligned}$$

From Part (a) of Lemma 2, we have the following equation in exchangeable case:

$$MR_n^1(t) = \sum_{i=1}^{n} p_i \sum_{j=n-i+1}^{n} (-1)^{j-n+i-1} \binom{n}{j} \binom{j-1}{n-i} H_{j,n}^1(t)$$

where

$$H_{j,n}(t) = E(T_{1:j} - t|T_{1:n} > t) = \int_0^\infty \frac{\bar{F}(\overbrace{t+x, \ldots, t+x}^{j}, \overbrace{t, \ldots, t}^{n-j})}{\bar{F}(t, \ldots, t)} dx$$

and $\bar{F}(x_1, \ldots, x_n) = P(T_1 > x_1, \ldots, T_n > x_n)$ is the joint survival function of T_i's. Since $\sum_{i=1}^{n} p_i = 1$ it is easy to see that

$$\begin{aligned} H_n^1(t) &= E(T_{1:n} - t|T_{1:n} > t) \le MR_n^1(t) \\ &= E(T - t|T_{1:n} > t) \le H_n^n(t) = E(T_{n:n} - t|T_{1:n} > t). \end{aligned}$$

Now suppose that there exists $1 < k \le n$ such that $p_i = 0$ for $i = 1, 2, \ldots, k-1$. $(n-k+1)$-out-of-n structure and consecutive-k-out-of-n:F system are well known examples of the systems that have this property. For such a system we define mean residual life as follow:

$$\begin{aligned} MR_n^{r,k}(t) &= E(T - t|T_{r:n} > t) = \sum_{i=k}^{n} p_i E(T_{i:n} - t|T_{r:n} > t) \\ &= \sum_{i=k}^{n} p_i M_n^{r,i}(t), 1 \le r \le k \le n. \end{aligned}$$

In view of Theorem 5, we note that $MR_n^{r,k}(t)$ is a decreasing function of r, $r = 1, 2, \ldots, k$. Hence we have

$$MR_n^{k,k}(t) = E(T - t|T_{k:n} > t) \leq MR_n^{r,k}(t) = E(T - t|T_{r:n} > t) \leq MR_n^{1,k}(t) = E(T - t|T_{1:n} > t).$$

4. Coclusion Remarks

We obtained in Section 2, several results about the two mean residual life functions of a $(n - k + 1)$-out-of-n system

$$H_n^k(t) = E(T_{k:n} - t|T_{1:n} > t), \;\; M_n^{r,k}(t) = E(T_{k:n} - t|T_{r:n} > t), 1 \leq r \leq k \leq n$$

when the lifetimes of the system components are independent random variables but not necessary identically distributed and when the joint distribution of the components lifetimes is exchangeable. A sufficient condition for $H_n^k(t)$ to be a decreasing(increasing) function of t, in nonidentical case and for $H_n^1(t)$ in exchangeable case is given. We have shown in nonidentical case that when the components of the system have increasing failure rate, $M_n^{r,k}(t)$ is decreasing in time. In both above mentioned cases it is shown that $M_n^{k,k}(t) \leq M_n^{r,k}(t) \leq H_n^k(t)$. Also it is shown in both cases that $M_n^{r,k}(t)$ is a decreasing function of r. We showed that $H_{n-1}^{k-1}(t) \leq H_n^k(t) \leq H_{n-1}^k(t)$, if T_n is independent of $T_1, \ldots, T_{n-1}$. Using the concept of signature introduced by Samaniego (1985), the MRL function of a coherent structure with exchangeable components is considered in Section 3.

References

[1] Asadi, M. and Goliforushani, S.(2008) On the mean residual life function of coherent systems. IEEE Transactions on Reliability, 57, 574-580.

[2] Barlow, R.E and Proschan, F.(1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart and Wineston, New York.

[3] David, H. and Nagaraja, H.N.(2003). Order Statistics. 3rd edition. John Wiley and Sons.

[4] Kochar, S.C., Mukherjee, H. and Samaniego, F.J.(1999). The 'signature' of a coherent system and its application to comparisons among systems. Naval Research Logistics, 46, 507-523.

[5] Navarro, J., Ruiz, J.M. and Sandoval, C.J.(2005). A note on comparisons among coherent systems with dependent components using signature. Statistics and Probability Letters, 72, 179-185.

[6] Navarro, J., Samaniego, F.J., Balakrishnan, N. and Bhattacharya, D.(2008). On the application and extension of system signatures in engineering reliability. Naval Research Logistics, 55, 313-327.

[7] Khanjari, S.M.(2008). Mean past and mean residual life functions of a parallel system with nonidentical components. Communications in Statistics: Theory and Methods, 37, 1134-1145.

[8] Samaniego, F.J. (1985). On the closure of the IFR class under formation of coherent systems. IEEE Transactions on Reliability, 34, 69-72.

A Comparative Study on Tests Concerning a Parameter Vector

M. Mahdizadeh

Department of Statistics, Ferdowsi University of Mashhad

Testing hypothesis about a parameter vector is an important statistical problem received much attention in parametric and nonparametric settings. In this work, we investigate connection between some types of directional tests and establish that two elements of the above class of tests are equivalent.

Keywords: Directional tests; Testing hypothesis, Quadratic form.

1. Introduction

Suppose that $H_0 : \boldsymbol{\beta} = \mathbf{0}$ denotes the hypothesis of interest, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is a k dimensional vector and $\mathbf{0}$ is the zero vector. The parameters can represent, for example, different types of treatment. An estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_k)'$ of $\boldsymbol{\beta}$ can be obtained by maximizing an objective function $H_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n})$, where $\boldsymbol{\lambda}$ is a vector of nuisance parameters and $\boldsymbol{x_n}$ denotes data provided by n independent samples. Some choices for objective function are likelihood, partial likelihood, etc. Assuming some regularity conditions, $\hat{\boldsymbol{\beta}}$ can be shown to be consistent and asymptotically normal under $H_0 : \boldsymbol{\beta} = \mathbf{0}$, i.e. $\hat{\boldsymbol{\beta}} \xrightarrow{p} 0$ and

$$\sqrt{n}\hat{\boldsymbol{\beta}} \xrightarrow{d} (0, \Sigma), \tag{1}$$

as $n \to \infty$, where Σ is a covariance matrix with rank k. There are situations that we want to have a test to be sensitive to alternatives of the form $\boldsymbol{\beta} = \boldsymbol{d}\beta^*$, where $\boldsymbol{d} = (d_1, \ldots, d_k)'$ is known and the scaler $\beta^* \neq 0$ satisfies

$$\frac{\beta_1}{d_1} = \ldots = \frac{\beta_k}{d_k} = \beta^*. \tag{2}$$

For example, $\boldsymbol{d} = (1, \ldots, 1)'$ represents a directional test in which the elements of $\boldsymbol{\beta}$ are equal. Also, $d_1 \leq \ldots \leq d_k$ implies a trend among the elements of $\boldsymbol{\beta}$.

A systematic approach to construct a directional test is adding (2) (as a constraint) to the objective function. This is equivalent to maximizing the function $H_n(\boldsymbol{d}\beta^*, \boldsymbol{\lambda}, \boldsymbol{x_n})$ with respect to β^*. Let $\hat{\beta}^*$ maximizes $H_n(\boldsymbol{d}\beta^*, \boldsymbol{\lambda}, \boldsymbol{x_n})$ and consider directional test statistic $T_1(\boldsymbol{d}) = n(\hat{\beta}^*)^2/\hat{\sigma}^2$, where $\hat{\sigma}^2$ is a consistent estimator of

the variance of $\sqrt{n}\hat{\beta}^*$. From (1), it follows that $T_1(\boldsymbol{d})$ is asymptotically distributed as χ^2_1 under H_0. The test is invariant to changes to $\boldsymbol{d}$. When the objective function is a likelihood, $T_1(\boldsymbol{d})$ is Wald test, and Thus asymptotically equivalent to the corresponding likelihood ratio and score tests of H_0 (Cox and Hinkley, 1974).

Another approach for obtaining a directional test is forming a linear combination of the components of $\hat{\boldsymbol{\beta}}$ obtained from the unconstrained objective function $H_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n})$. That is to define the directional test statistic $T_2(\boldsymbol{c}) = n(\boldsymbol{c}'\hat{\boldsymbol{\beta}})^2/(\boldsymbol{c}'\hat{\Sigma}\boldsymbol{c})$, where $\boldsymbol{c}$ is $k \times 1$ vector. Again using (1) it follows that $T_2(\boldsymbol{c})$ asymptotically has χ^2_1 distribution under H_0, and it is invariant to scale changes to $\boldsymbol{c}$.

Examples of tests are numerous in the literature. A commonly used test is the Cochran-Armitage trend test for binary responses which arises as a score test from logistic and other regression models (Tarone and Gart, 1980) in which there is a single covariate with values $d_1, \ldots, d_k$ corresponding to k groups. A trend test similar to the constrained test for detecting a dose-response was discussed by Tukey et al. (1985).

To motivate this study, we mention two important examples of previous studies on this issue. The first one is the method proposed by Wei et al. (1989) for analyzing multivariate survival time data using the marginal distributions of an individual's K survival times. Let β_k be the log-risk ratio between two treatment groups for the kth of K survival times. The objective function can be expressed as

$$H_n(\boldsymbol{\beta}, \boldsymbol{x_n}) = \prod_k \prod_i \Big(\frac{\exp(\beta_k z_i)}{\sum_j Y_{kj}(x_{ki}) \exp(\beta_k z_j)} \Big)^{\delta_{ki}}, \tag{3}$$

where z_i denotes treatment indicator of subject i $(i = 1, \ldots, n)$, x_{kj} is the observed portion of the kth event for the ith subject, δ_{kj} is the censoring indicator for the kth event and ith subject, and $Y_{kj}(x_{ki})$ indicates if subject j is in the risk set at time x_{ki} for the kth event. That is $\beta_1, \ldots, \beta_k = \beta^*$. Thus we obtain the model

$$H_n(\mathbf{1}\beta^*, \boldsymbol{x_n}) = \prod_k \prod_i \Big(\frac{\exp(\beta^* z_i)}{\sum_j Y_{kj}(x_{ki}) \exp(\beta^* z_j)} \Big)^{\delta_{ki}}, \tag{4}$$

where β^* is the common treatment effect. Lin (1994) proposed a directional test of H_0 based on (4). The estimate of common risk ratio $\hat{\beta}^*$ can be obtained by maximizing (4), and then one can construct the test $T_1(\boldsymbol{d})$. Similarly, the test $T_2(\boldsymbol{c})$ can be obtained using the estimates by maximizing the model (3). Wei et al. (1989) proposed such a test by taking $\boldsymbol{c} = (\hat{\Sigma}^{-1}\mathbf{1})/(\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1})$, where $\hat{\Sigma}$ is a consistent estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$.

As a second example, an objective function with the same form as (3) arises in comparing two treatment groups with respect to survival time when a stratified proportional hazard model is used (Kalbfleisch and Prentice, 2002), with strati-

fication based on the levels of a categorical covariate. In this situation, the most common method of testing for a treatment difference is a $T_1(\boldsymbol{d})$ type test that assumes a common treatment relative risk across strata.

Despite the extensive use of these directional tests, little effort has been made on investigating relationship between them. To this end, we consider the following sequence of alternatives to H_0

$$H_A^n : \boldsymbol{\beta} = \frac{\boldsymbol{\mu}}{\sqrt{n}}, \tag{5}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ is a constant vector and n is the sample size. We assume that the objective function leads to a consistent and asymptotically normal estimate of $\boldsymbol{\beta}$. Section 2 is devoted to introducing some notations. The main result of the paper is given in Section 3. We end in Section 4 with a summary.

2. Notations

In this section we introduce some notations which facilitate presentation of the proof in the next section. If (1) holds, then under the sequence of alternatives in (5),

$$\sqrt{n}\hat{\boldsymbol{\beta}} \xrightarrow{d} (\boldsymbol{\mu}, \Sigma)$$

as $n \to \infty$. In general, $\Sigma = {I^{\beta\beta}}^{-1} V^{\beta\beta} {I^{\beta\beta}}^{-1}$ (see Rao, 1973), where

$$I^{\beta\beta} = I_{\beta\beta} - I_{\beta\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\beta}$$

and

$$V^{\beta\beta} = V_{\beta\beta} - I_{\beta\lambda} I_{\lambda\lambda}^{-1} V_{\lambda\beta} - V_{\beta\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\beta} + I_{\beta\lambda} I_{\lambda\lambda}^{-1} V_{\lambda\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\beta}.$$

In the above expressions, $V_{\beta\beta}$, $V_{\beta\lambda}$, $V_{\lambda\beta}$ and $V_{\lambda\lambda}$ are the elements of the covariance matrix V of the score vector from $H(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n})$ and $I_{\beta\beta}$, $I_{\beta\lambda}$, $I_{\lambda\beta}$ and $I_{\lambda\lambda}$ are the components of the probability limits I_D of the negative of the second derivative of $H(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n})/n$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, where

$$V = \begin{pmatrix} V_{\beta\beta} & V_{\beta\lambda} \\ V_{\lambda\beta} & V_{\lambda\lambda} \end{pmatrix}$$

and

$$I_D = \begin{pmatrix} I_{\beta\beta} & I_{\beta\lambda} \\ I_{\lambda\beta} & I_{\lambda\lambda} \end{pmatrix}.$$

When H is a likelihood function, I_D is the corresponding information matrix.

3. Theoretical results

Assume the distribution of $\boldsymbol{x_n}$ depends on parameters $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ and the objective function $H_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n})$ is continuous and has first and second derivatives. Let $\boldsymbol{\beta}_0$ and $\boldsymbol{\lambda}_0$ maximizes $H(\boldsymbol{\beta}, \boldsymbol{\lambda}) = E(H_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n}))$ under the sequence of alternatives H_A^n in (5), so that $H(\boldsymbol{\beta}, \boldsymbol{\lambda})$ is maximized at $(\mathbf{0}, \boldsymbol{\lambda}_0)$ as $n \to \infty$. Define

$$U_\beta(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{\partial}{\partial \boldsymbol{\beta}} H_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n})$$

and

$$U_\lambda(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{\partial}{\partial \boldsymbol{\lambda}} H_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n})$$

and assume that

$$\frac{1}{\sqrt{n}} \begin{pmatrix} U_\beta(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0) \\ U_\lambda(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0) \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, V) \tag{6}$$

under H_A^n, where V is a covariance matrix and $V_{\beta\beta}$, $V_{\beta\lambda}$, $V_{\lambda\beta}$ and $V_{\lambda\lambda}$ are the corresponding covariance submatrices for parameters $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ in V. From (6) we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} (I_{\beta\beta} - I_{\beta\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\beta})^{-1} \big(U_\beta(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0) - I_{\beta\lambda} I_{\lambda\lambda}^{-1} U_\lambda(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0) \big) + o_p(1), \tag{7}$$

where $I_{\beta\beta}$, $I_{\beta\lambda}$, $I_{\lambda\beta}$ and $I_{\lambda\lambda}$ are defined as the probability limits of the negative of the second derivative of $H_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{x_n})/n$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. Therefore $\sqrt{n}\hat{\boldsymbol{\beta}} \xrightarrow{d} (\boldsymbol{\mu}, \Sigma)$.

Consider the model $H_n(\boldsymbol{d}\beta^*, \boldsymbol{\lambda}, \boldsymbol{x_n})$ which is obtained by setting $\boldsymbol{\beta} = \boldsymbol{d}\beta^*$. $H(\boldsymbol{d}\beta^*, \boldsymbol{\lambda})$ is maximized at $(\beta_0^*, \boldsymbol{\lambda}_0^*)$ under the sequence of alternatives in (5). Since $\beta_0^* \to 0$ as $n \to \infty$, it follows that $H(\boldsymbol{d}\beta^*, \boldsymbol{\lambda})$ is maximized at $(0, \boldsymbol{\lambda}_0)$ as $n \to \infty$.

Let $\beta_0^* = (\boldsymbol{d}' I^{\beta\beta} \boldsymbol{\beta}_0)/(\boldsymbol{d}' I_{\beta\beta} \boldsymbol{d})$. Since β_0^* tends to zero as $n \to \infty$

$$\frac{1}{\sqrt{n}} \begin{pmatrix} U_\beta(\boldsymbol{d}\beta_0^*, \boldsymbol{\lambda}_0) \\ U_\lambda(\boldsymbol{d}\beta_0^*, \boldsymbol{\lambda}_0) \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, V).$$

Let $(\hat{\beta}^*, \hat{\boldsymbol{\lambda}}^*)$ be the solution of $\boldsymbol{d}' U_\beta(\boldsymbol{d}\beta^*, \boldsymbol{\lambda}) = 0$ and $U_\lambda(\boldsymbol{d}\beta^*, \boldsymbol{\lambda}) = 0$. A Taylor series expansion of $(\boldsymbol{d}' U_\beta(\boldsymbol{d}\beta_0^*, \boldsymbol{\lambda}_0), U_\lambda(\boldsymbol{d}\beta_0^*, \boldsymbol{\lambda}_0))'$ at $(\hat{\beta}^*, \hat{\boldsymbol{\lambda}}^*)/\sqrt{n}$ gives

$$\sqrt{n}(\hat{\beta}^* - \beta_0^*) = \frac{1}{\sqrt{n}} \frac{\boldsymbol{d}' (U_\beta(\boldsymbol{d}\beta_0^*, \boldsymbol{\lambda}_0) - I_{\beta\lambda} I_{\lambda\lambda}^{-1} U_\lambda(\boldsymbol{d}\beta_0^*, \boldsymbol{\lambda}_0))}{\boldsymbol{d}' (I_{\beta\beta} - I_{\beta\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\beta}) \boldsymbol{d}} + o_p(1). \tag{8}$$

And it follows that

$$\sqrt{n}(\hat{\beta}^* - \beta_0^*) \xrightarrow{d} N(0, \frac{\boldsymbol{d}' V^{\beta\beta} \boldsymbol{d}}{(\boldsymbol{d}' I^{\beta\beta} \boldsymbol{d})^2}).$$

If we consider $T_2(\boldsymbol{c})$ with $c = (I^{\beta\beta}\boldsymbol{d})/(\boldsymbol{d}'I^{\beta\beta}\boldsymbol{d})$, then from (7) and (8) we have

$$\sqrt{n}(\hat{\beta}^* - \boldsymbol{c}'\hat{\boldsymbol{\beta}}) \xrightarrow{p} 0.$$

Hence we conclude $T_1(\boldsymbol{d}) - T_2(\boldsymbol{c}) \xrightarrow{p} 0$ as $n \to \infty$.

4. Conclusion

Testing hypothesis about a parameter vector is an important statistical problem. In this article we investigated relationship between some elements of the class of directional tests and showed there exists an equivalence property. For a given $T_1(\boldsymbol{d})$, consider $T_2(\boldsymbol{c})$, for which $\boldsymbol{c}$ is proportional to $I^{\beta\beta}\boldsymbol{d}$. It was shown that as the sample size increases, the probability that these two tests will differ more than any arbitrarily small value goes to zero.

References

1. Cox, D. R. and Hinkley, D. V. (1974). *Theoretical statistics.* Chapman & Hall, London.
2. Kalbfleisch, J. D., and Prentice, R. L. (2002). *The statistical analysis of failure time data,* 2nd edn. John Wiley and Sons, Hoboken, NJ.
3. Lin, D. Y. (1994). Cox regression of multivariate failure time data: the marginal approach. *Statist. Med.* **13,** 2233 - 2247.
4. Rao, C. R. (1973). *Linear Statistical Inference and Its Applications,* 2nd ed. Wiley, New York.
5. Tarone, R. E. and Gart, J. J. (1980). On the robustness of combined tests for trends in proportions. *J. Amer. Statist. Assoc.* **75,** 110 - 116.
6. Tukey, J. W., Ciminera, J. L., and Heyse, J. F. (1985). Testing the statistical certainty of a response to increasing dose of a drug. *Biometrics* **41**, 295 -301.
7. Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84,** 1065 - 1073.

Modified Control Limits For The Mean Of Autoregressive Process

F. Misagh[1], H. Namdar[1], G. H. Yari[2] and A. Hosseini[1]

[1] *Department of Mathematics and Statistics, Islamic Azad University- Tabriz branch*

[2] *School of Mathematics, Iran University of Science and Technology*

[3] *Workers' House of Tabriz, University of Applied Science and Technology*

This paper describes the application of six-sigma control limits in serially correlated data. Performance of traditional control limits has been investigated in autocorrelated data. Modified control limits have established via longevous variation of process that guarantee the certain inferences and reliable consequences. The behavior of traditional and modified control limits was evaluated for independent and correlated data. Average Run Length criterion was used to discover a shift in the mean of process and to compare the efficiency of control limits based on longevous and pseudo variation.

Keywords: Quality Control; Uncertain Process; Autoregressive; Average Run Length.

1. Introduction

Statistical process control, SPC, refers to a number of different methods for monitoring and assessing the quality of manufactured goods. Combined with methods from the Design of Experiments, SPC is used in programs that define, measure, analyze, improve, and control development and production processes. These programs are often implemented using Design for Six Sigma methodologies.

Briefly, a control chart is a graphical method for detecting if the underlying distribution of variation of some measurable characteristic of the product seems to have undergone a shift. Such a shift likely reflects a subtle drift or change to the desired manufacturing process that needs to be corrected in order to maintain good quality output, Shewhart, W. (1931).

In traditional SPC, it is assumed that the observations are independently and normally distributed, but independence of data may not be established practically. Our goal is to determine the impression of correlation on the efficiency of traditional control limits for the mean by using Average Run Length, ARL, criterion.

ARL criterion has been of interest in the literature of SPC, Goldsmith, P. L. and H. Whitfield (1961).

2. Average Run Length

Suppose X_t denotes independent observations of a continuous quality characteristics with normal distribution, in other words:

$$X_t \sim N(\mu_t, \sigma_X^2) \quad \text{for } t \in \mathcal{N},$$

where $\mathcal{N}$ denotes the set of natural numbers. We assume that a harmony causes a shift in the mean at an unknown time T,

$$\mu_t = \begin{cases} \mu & \text{for } t < T, \\ \mu + \delta\sigma_X & \text{for } t > T. \end{cases}$$

In SPC, the six-sigma control limits are the form of $\mu \pm 3\sigma_X$. After a shift in the mean, an observation may not fall between Lower Control Limit, LCL, and Upper Control Limit, UCL. We denote the probability of falling an observation within control limits by $P(\delta)$,

$$\begin{aligned} P(\delta) &= P(\mu - 3\sigma_X \leq X_t \leq \mu - 3\sigma_X) \\ &= \Phi(\delta + 3) - \Phi(\delta - 3). \end{aligned}$$

where $\Phi(\cdot)$ is the Cumulative Distribution Function of the standard normal distribution. Suppose N denotes the number of observations until the first out-of-control one; N has Geometric distribution with parameter $P(\delta)$ and The ARL is the average of N, i.e:

$$ARL(\delta) = \frac{1}{1 - P(\delta)}$$

The ARL curve of a six-sigma and five-sigma control limits are represented in Fig. 1. For $\delta = 0$, ARL is maximum and ARL(δ) is decreasing function of δ.

2.1. *Evaluating Autoregressive process data in traditional limits*

Suppose Y_t denotes the AR(1) time series, i.e:

$$Y_t - \mu = \theta(Y_{t-1} - \mu) + \xi_t, \tag{1}$$

where ξ_t are pairwise independently and identically distributed, i.i.d, random variables of $N(0, \sigma_\xi^2)$; it is easy to prove that $\sigma_Y^2 = \sigma_\xi^2/(1 - \theta)$.

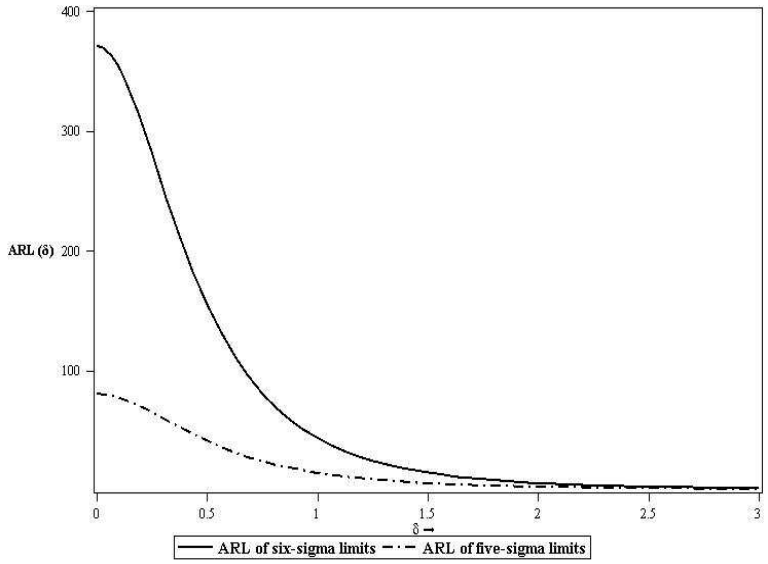

Fig. 1. ARL curve of a traditional control limits.

Suppose $Y_1, \cdots Y_n$ are observations of AR(1) process. In traditional SPC, central line of control chart is determined by sample mean $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and standard deviation is substituted by $\overline{MR}/d(2)$, where

$$\overline{MR} = \frac{1}{n-1}\sum_{i=2}^{n} MR_i,\ MR_i = |Y_i - Y_{i-1}|$$

and $d(2) \simeq 1.13$ is commonly used in the literature of SPC, Montgomery, D. C. (1996).

To evaluate the behavior of AR(1) process in SPC, we simulate a sequence of size 1000 from model (1) with $\mu = 0,\ \sigma_\xi^2 = 1$ and five values of θ = - 0.8; - 0.4; 0; 0.5; 0.9; This simulation is repeated 10,000 times and the results is given in table 1. Values of σ_Y calculated by the relation between θ and σ_ξ^2, but $E[\overline{MR}/d(2)]$ and $ARL(\delta)$ are the average of the values generated in every step of simulation process.

In table 1, comparing σ_Y, real or longevous variation, and $E[\overline{MR}/d(2)]$, pseudo or short term variation, shows over-estimation and under-estimation of AR(1) process variance for negative θ and positive θ respectively.

For negative θ, the width of control limits is so large for negative θ and values

Table 1. Evaluating traditional six-sigma control limits in correlated data.

θ	σ_Y	$E[\overline{MR}/d(2)]$	LCL	UCL	ARL(0)	ARL(1)
-0.8	2.7778	2.6448	-7.9344	7.9344	32582.98	1186.43
-0.4	1.1905	1.2241	-3.6723	3.6723	3451.85	219.18
0	1.0000	1.0012	-3.0036	3.0036	353.93	61.22
0.5	1.3334	0.7985	-2.3955	2.3955	41.99	13.52
0.9	2.2942	0.6998	-2.0994	2.0994	9.23	3.46

of two last columns show puzzling and unexpected number of observations until first out-of-control case. Vise versa, for positive θ, nearly all of observations are recognized out-of-control. Therefore, using traditional control limits to evaluating correlated data may cause serious problems in the interpretation of SPC design. A suggestion to overpowering these problems is to construct LCL and UCL based on σ_Y instead of $\overline{MR}/d(2)$.

3. Modified control limits for autocorrelated data

In this section, modified control limits will be established in base of longevous variation of process. Subsequently, the new limits will be evaluated using ARL criterion.

Suppose Y_t is an AR(1) time series with parameter θ, i.e:

$$Y_t - \mu_t = \theta(Y_{t-1} - \mu_{t-1}) + \xi_t, \quad for\, t \in \mathbb{Z}$$

where $|\theta| < 1$ and ξ_t is an i.i.d sequence of $N(0, \sigma_\xi^2)$ distribution. The value of Y_t is exactly determined by Y_{t-1} and so, the AR(1) model is known as Morkov process, Box, G. E. P. and Jenkins, G. M. (1976).

Suppose $Y_1, \cdots, Y_n$ are the realizations of AR(1) process; similar to i.i.d data, We estimate the central line of control chart through $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. LCL and UCL are determined by $\overline{Y} \pm 3\widehat{\sigma}_Y$. In practice, the values of σ_Y and σ_ξ can be estimated by the following equalities:

$$\widehat{\sigma}_Y = C_0 = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \overline{Y})^2, \qquad \widehat{\sigma}_\xi = \frac{C_0}{1+\widehat{\theta}^2}$$

$$r = \frac{\sum_{t=1}^{n-1}(Y_t - \overline{Y})(Y_{t-1} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}, \qquad r = \frac{-\widehat{\theta}}{1+\widehat{\theta}^2}$$

where r denotes the autocorrelation estimator.

In the remaining of this section, we assume values of θ and σ_Y are known; Otherwise it is easy to estimate the unknown parameters from a realization of process that is accessible in practical situations.

3.1. *Evaluating Autoregressive data in modified control limits*

Suppose a phenomenon causes a shift in the mean from μ to $\mu+\delta\sigma_Y$ at an unknown instant of T. Therefore, the model (1) can be rewritten as:

$$(Y_t - \mu) - \delta\sigma_Y = \theta(Y_{t-1} - \mu) - \theta\delta\sigma_Y + \xi_t$$

For $Y_{t-1} = s$, the value of Y_t is equal to

$$\begin{aligned} v &= \mu + \delta\sigma_Y + \theta(s - (\mu + \delta\sigma_Y)) + \xi_t \\ &= \theta s + (1-\theta)\mu + (1-\theta)\delta\sigma_Y + \xi_t. \end{aligned}$$

The Run Length equals one if the value of v falls out of control limits. Otherwise, it is one plus Run Length of AR(1) process with initiation point v. We denote the ARL of AR(1) process by $L_\theta(\delta, s)$ that s is the beginning state of process. Referring to discussion above, we have:

$$\begin{aligned} L_\theta(\delta, s) &= 1 + \int_{\xi|\mu-3\sigma_Y \le v \le \mu+3\sigma_Y} L_\theta(\delta, v) f(\xi) d\xi \\ &= 1 + \int_{\mu-3\sigma_Y}^{\mu+3\sigma_Y} L_\theta(\delta, v) f(v - \theta s - (1-\theta)(\mu + \delta\sigma_Y)) dv. \end{aligned}$$

where $f(\xi)$ is the probability density function of normal random variable ξ. With regard to Fredholm Integral, Arfken, G. (1985), there is a recurrent sequence of functions $\{L_0, L_1, L_3, \cdots\}$ so that $lim_{k\to\infty} L_k(\delta, s) = L_\theta(\delta, s)$ and

$$\begin{aligned} L_i(\delta, s) &= 1 + \int_{\mu-3\sigma_Y}^{\mu+3\sigma_Y} L_{i-1}(\delta, v) f(v - \theta s - (1-\theta)(\mu + \delta\sigma_Y)) dv \\ &= 1 + E[L_{i-1}(\delta, V) I_{(\mu-3\sigma_Y, \mu+3\sigma_Y)}(V)]. \end{aligned}$$

where V is a random variable with $N(\eta, 1)$ distribution and $\eta = \theta s + (1-\theta)(\mu + \delta\sigma_Y)$. L_0 is an arbitrary continuous function on $(\mu - 3\sigma_Y, \mu + 3\sigma_Y)$; for example, $L_0 = 1$ can be a suitable choice.

Solving or simulating Fredholm integral to reach the value of $L_\theta(\delta, s)$ is hard and complicated. To get to this goal, we choose an abridged way by simulating

AR(1) process until one of v values mentioned above falls outside of control limits. According to definition of ARL criterion, The number of v's up to the first uncertain one will set as the ARL value. Assuming $s = 0$,we repeat the simulation 10,000 times per every θ; relevant values of $L_\theta(\delta, 0)$ are the mean of ARL values generated in every step of simulation that is presented in table below. Compatible with i.i.d case, the quantities of $L_\theta(\delta, 0)$ have been denoted by $ARL(\delta)$.

Table 2. Evaluating modified control limits in autocorrelated data.

θ	ARL(0)	ARL(0.5)	ARL(1)	ARL(1.5)	ARL(2)	ARL(2.5)
-0.8	378.64	139.03	59.66	15.14	6.09	1.13
-0.4	368.95	159.29	56.82	12.59	8.05	2.01
0	372.65	153.61	45.76	10.54	8.98	2.76
0.5	381.38	187.08	61.77	12.68	13.09	1.99
0.9	375.32	260.45	105.49	42.77	32.03	15.22

Despite of table 1, ARL values in table 2 show logical and desired values for autocorrelated data; So, using longevous variation instead of pseudo variation, improves traditional control limits to take advantage in AR(1) data.

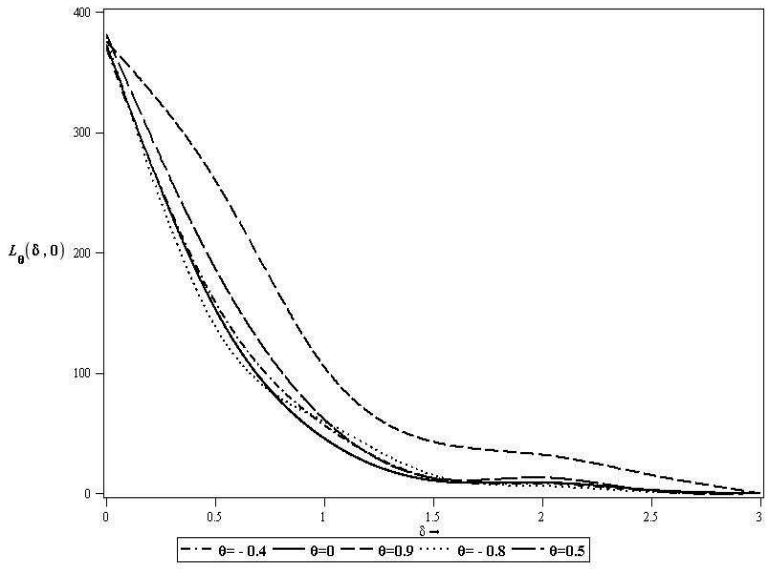

Fig. 2. Curve of $L_\theta(\delta, 0)$ for modified control limits.

For $\theta = 0$, the values of $L_\theta(\delta, 0)$ is the same as ARL(δ) values depicted in table 1; We know that for $\theta = 0$, AR(1) process generates i.i.d normally distributed random variables, Box, G. E. P. and Jenkins, G. M. (1976). Comparing ARL values of different θ shows that for positive values of θ, the modified control limits has low sensitivity to explore the shift in the mean or AR(1) process. ARL values for $\theta = 0.9$ are too large comparative to others.

For $\theta = -0.8$ and $\theta = 0$, $L_{\theta=-0.8}(\delta = 0.5, 0) < L_{\theta=0}(\delta = 0.5, 0)$ indicates that for δ close to 0.5, modified control limits has high sensitivity to detecting change in the mean of process.

For other values of θ, except for $\theta = 0.9$, modified control limits is powerful as well as Traditional control limits in discovering behavior of uncertain process.

The functions $L_\theta(\delta, 0)$ are curved in Fig. 2; The curve for $\theta = 0$ is so similar to curve considered in Fig. 1 and all curves have decreasing trend and tend to zero. Except the similar origin latitude with other curves, $L_{\theta=o.9}(\delta, 0)$ is very elevated and more different from independent case curve. Other curves are almost the same as i.i.d case, except for in the region nearby $\delta = 0.5$; average run length for $\theta = -0.8$ is less than the others. Therefore, it can be concluded that in detecting small changes of mean; the more negative θ, the more efficient the modified control limits will be.

4. Conclusion

In the previous sections, the behavior of independent and autocorrelated data in traditional and modified control limits have been examined. Using Traditional control limits may cause miss-understandings of production process. Traditional control limits has been adopted using longevous variation instead of pseudo variation.

In conclusion, if one is interested in checking Autocorrelated process for changes in the mean using modified control limits, for negative and not too large positive θ, the ARL criterion shows enough efficiency comparable to a traditional control limits for independent observations.

References

1. G. Arfken, *Mathematical Methods for Physicists.* (Academic Press, 3rd ed, p. 865, Orlando, FL, 1985).
2. G. E. P. Box, and G. M. Jenkins, *Time Series Analysis, Forecasting and Control.* (revised edition, Holden-Day, San Francisco, 1976).

3. P. L. Goldsmith, H. Whitfield, *Average run length in cumulative limits quality control schemes.* (Technometrics, 3(1), 11(20), 1961).
4. D. C. Montgomery, *Introduction to Statistical Quality Control, third.* (third edition, John Wiley & Sons, New York., 1996).
5. W. Shewhart, *Economic Control of Quality of Manufactured Product.* (D. van Nostrand Company, Inc., Toronto, 1931).

Modified Regression Estimators for Rare Populations

M. Moradi and M. Salehi

Department of Science, Razi University, Kermanshah

When we have an auxiliary variable that is correlated to the response variable the regression estimator is an efficient. In a rare population having a lot of zero values, regression estimator or ratio estimator is undefined for those samples containing only non-rare subpopulation (zero values) in simple random sampling. In this paper, we introduce modified regression estimators and their variance estimators for sampling designs which are suitable to rare populations, such as general inverse sampling and inverse sampling with unequal selection probabilities. Also we conduct a simulation study on a real rare population. The simulation results show that the modified regression estimators in both sampling designs are more efficient than conventional estimator in simple random sampling and the modified regression estimator in inverse sampling with unequal selection probabilities is more efficient than others.

Keywords: General Inverse sampling; Murthy's estimator; Unequal selection probabilities ; Auxiliary variables.

1. Introduction

Usually rare populations contain a lot of non-rare units and a few rare ones. For example, when we estimate number of rare animals (or plants) in a wide area, we partition the area to a large number of quadrates, the variable of each sampling unit (here each quadrat is a sampling unit) is the number of interested animals (or plants) in such a unit. Obviously if we implement simple random sampling design, it is likely the most of selected sample units to be from non-rare units and the sample set has not enough information about rare units. In such cases, if all sample units are selected from non-rare subpopulation, the conventional estimator of rare subpopulation's total will be equal to zero. And other estimators such as ratio and regression will be undefined (Moradi *et al.*, 2007). [1]

In survey sampling of a rare population one probability sampling scheme is desired which intend to select more units from the rare subpopulation. Inverse sampling design is one of such sampling designs that was first proposed by Haldane (1945) in which one continues sampling until a pre-determined number of rare events of interest is observed. [2]

One deficiency, however is that the final sample size is not fixed, which makes it

difficult to plan budgets and survey logistics. To get round this problem Salehi and Seber (2004) proposed a general inverse sampling design. In general inverse sampling design, sample sizes n_0 and n_1 are determined based on a minimum and a maximum budget (or time). First, an initial sample of size n_0 is taken. If the initial sample has the pre-determined number of rare events the sampling would be stopped. Otherwise the sampling is sequentially continued until to achieve the pre-determined number of events or reach the sample size n_1. Salehi and Seber (2004) also derived unbiased Murthy's estimator for their sampling design. [3]
Greco and Naddeo (2007) introduced inverse sampling with unequal selection probabilities and they proposed an unbiased population mean estimator for their sampling design. [4]
If selection probabilities is approximately proportional to the variable of study in inverse sampling with unequal selection probabilities, we get a predetermined number of rare units with a smaller final sample size than when we use inverse sampling design to get such number of rare units.
Additional to improving the sampling design using auxiliary variables, we can utilize auxiliary variables to improve estimators. For example, when we have an auxiliary variable that is correlated with the response variable, we can use ratio estimator or regression estimator to improve the efficiency. Moradi *et al.* (2007) proposed a ratio estimator using Murthy's estimator for general inverse sampling design.
In this paper, we introduce a modified regression estimators for general inverse sampling design and inverse sampling with unequal selection probabilities for which variance of modified regression estimators as well as variance estimators will be developed . Using a simulation study on a real population Arsenic (AS) contaminant in a region of Kurdistan province, we show modified regression estimators are more efficient than their counterparts.

2. Modified regression estimator

In simple random sampling the regression estimator is defined as following

$$\bar{y}_l = \bar{y} + b(\bar{X}_N - \bar{x}).$$

where $b = (\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y})/(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)$ is an estimator of $B = (\sum_{i=1}^{N} x_i y_i - N\bar{X}\bar{Y})/(\sum_{i=1}^{N} x_i^2 - N\bar{X}^2)$. It is obvious that parameters $\sum_{i=1}^{N} x_i y_i/N$, $\bar{X}$, $\bar{Y}$, and $\sum_{i=1}^{N} x_i^2/N$ are estimated by corresponding sample means $\sum_{i=1}^{n} x_i y_i/n$, $\bar{x}$, $\bar{y}$, and $\sum_{i=1}^{n} x_i^2/n$, respectively. The sample mean in simple random sampling has some desire properties like unbiasedness which is not true for inverse sampling designs. We therefore find other suitable estimators of parameters of B for introduced inverse sampling designs.

2.1. *Regression estimator in general inverse sampling*

In general inverse sampling we use Murthy's estimators $\hat{\bar{Y}}_M$ and $\hat{\bar{X}}_M$ to estimate $\bar{Y}$ and $\bar{X}$ respectively. Salehi and Seber (2004) calculated $\hat{\bar{Y}}_M$ for general inverse sampling design as following, [3]

$$\hat{\bar{Y}}_M = \begin{cases} \frac{\sum_{i=1}^{n_0} y_i}{n_0}, & n_s = n_0 \quad , \quad |\mathcal{S}_c| \geq m; \\ \hat{p}\bar{y}_c + (1-\hat{p})\bar{y}_{c'}, & n_0 < n_s \leq n_1 \quad , \quad |\mathcal{S}_c| = m; \\ \frac{\sum_{i=1}^{n_1} y_i}{n_1}, & n_s = n_1 \quad , \quad |\mathcal{S}_c| < m. \end{cases} \tag{1}$$

Estimator $\hat{\bar{X}}_M$ can be calculated similarly. Also we need to estimate $\sum_{i=1}^{N} x_i y_i$ and $\sum_{i=1}^{N} x_i^2$. Let $z_{1i} = x_i y_i$ and $z_{2i} = x_i^2$, using Murthy's estimator we have

$$\hat{\tau}_{z_1} = \sum_{i=1}^{n} \frac{P(\mathcal{S}|i)}{P(\mathcal{S})} z_{1i} = \sum_{i=1}^{n} \frac{P(\mathcal{S}|i)}{P(\mathcal{S})} x_i y_i, \quad \hat{\tau}_{z_2} = \sum_{i=1}^{n} \frac{P(\mathcal{S}|i)}{P(\mathcal{S})} z_{2i} = \sum_{i=1}^{n} \frac{P(\mathcal{S}|i)}{P(\mathcal{S})} x_i^2$$

Using (1) estimators $\hat{\tau}_{z_1}$ and $\hat{\tau}_{z_2}$ in general inverse sampling are given by

$$z_{j.M} = \begin{cases} \frac{N}{n_0} \sum_{i=1}^{n_0} z_{j\,i}, & n_s = n_0 \quad , \quad |\mathcal{S}_c| \geq m; \\ N(\hat{p}\bar{z}_{j\,c} + (1-\hat{p})\bar{z}_{j\,c'}), & n_0 < n_s \leq n_1 \quad , \quad |\mathcal{S}_c| = m; \\ \frac{N}{n_1} \sum_{i=1}^{n_1} z_{j\,i}, & n_s = n_1 \quad , \quad |\mathcal{S}_c| < m. \end{cases}$$

Then modified regression estimator in general inverse sampling design is given by

$$\bar{y}_{GI.l} = \hat{\bar{Y}}_M + b_M(\bar{X} - \hat{\bar{X}}_M)$$

where

$$b_M = \frac{\hat{\tau}_{z_1} - N\hat{\bar{X}}_M\hat{\bar{Y}}_M}{\hat{\tau}_{z_2} - N\hat{\bar{X}}_M^2}$$

An approximation method to calculate variance $V(\bar{y}_{Ml})$ for enough large sample size n, is that we assume $b_M \simeq B$ (similar to Cochran pp 194. [5]), then we have

$$\bar{y}_{GI.l} = \hat{\bar{Y}}_M - B\hat{\bar{X}}_M + B\bar{X} = \sum_{i=1}^{n} \frac{P(\mathcal{S}|i)}{NP(\mathcal{S})} y_i - B\sum_{i=1}^{n} \frac{P(\mathcal{S}|i)}{NP(\mathcal{S})} x_i + B\bar{X}$$

$$= \sum_{i=1}^{n} \frac{P(\mathcal{S}|i)}{NP(\mathcal{S})} (y_i - Bx_i) + B\bar{X} = \sum_{i=1}^{n} \frac{P(\mathcal{S}|i)}{NP(\mathcal{S})} V_i + B\bar{X}$$

where$V_i = y_i - Bx_i$. Then $V(\bar{y}_{GI.l}) = V(\hat{\bar{V}}_M)$ and using variance of Murthy's estimator we have

$$Var(\bar{y}_{GI.l}) \doteq \sum_{i=1}^{N}\sum_{j<i}^{N}(1 - \sum_{\mathcal{S}\ni i,j}\frac{P(\mathcal{S}|i)P(\mathcal{S}|j)}{P(\mathcal{S})})(\frac{V_i}{p_i} - \frac{V_j}{p_j})^2\frac{p_ip_j}{N^2},$$

and variance esimator is given by

$$\widehat{Var}(\bar{y}_{GI.l}) = \sum_{i=1}^{n}\sum_{j<i}^{n}(\frac{P(\mathcal{S}|i,j)}{P(\mathcal{S})} - \frac{P(\mathcal{S}|i)P(\mathcal{S}|j)}{P(\mathcal{S})^2})(\frac{\hat{V}_i}{p_i} - \frac{\hat{V}_j}{p_j})^2\frac{p_ip_j}{N^2}$$

where $\hat{V}_i = y_i - b_M x_i$.

Salehi and Seber (2004) calculated $\widehat{Var}(\hat{\bar{Y}}_M)$ in general inverse sampling, in order to calculate $\widehat{Var}(\bar{y}_{GI.l})$ in general inverse sampling we should replace $\hat{V}$ with y in their formula, then we have

$$\widehat{Var}(\bar{y}_{GI.l})$$
$$= \begin{cases} (\frac{1}{n_0} - \frac{1}{N})\frac{1}{n_0-1}\sum_{i=1}^{n_0}(\hat{V}_i - \bar{\hat{V}})^2, & n_s = n_0 \quad , \ |\mathcal{S}_c| \geq m; \\ \hat{p}^2(\frac{N(N-n+1)(nm-n-m)-N(n-2)}{(n-2)(m-1)})\frac{s^2_{c\hat{V}}}{m} + & n_0 < n_s < n_1 \ , \ |\mathcal{S}_c| = m; \\ N^2\widehat{var}[\hat{p}](\bar{\hat{V}}_c - \bar{\hat{V}}_{c'})^2 + (\frac{N(N-n+1)(n-m-1)}{(n-1)(n-2)})s^2_{c'\hat{V}} & \\ (\frac{1}{n_1} - \frac{1}{N})\frac{1}{n_1-1}\sum_{i=1}^{n_1}(\hat{V}_i - \bar{\hat{V}})^2, & n_s = n_1 \quad , \ |\mathcal{S}_c| < m. \end{cases}$$

2.2. *Regression estimator in inverse sampling with unequal selection probabilities*

Greco and Naddeo (2007) introduced inverse sampling with unequal selection probabilities. They consider a population of N units divided into two groups of N_1 and N_2 units, respectively. They let p_{1i} the selection probability of the ith unit in the first group and p_{2i} the selection probability of the ith unit in the second group, and they let $P = \sum_{i=1}^{N_1} p_{1i}$ the selection probability of the first group and $1 - P = \sum_{i=1}^{N_2} p_{2i}$ the selection probability of the second group.
Their sequential sampling with unequal selection probabilities continues until m units of the first group are observed in the sample, so that the sample size ν is a random variable. They proposed an unbiased estimator for total of population as

following

$$\tilde{T}_y = \frac{\hat{P}}{m}\sum_{i=1}^{m}\frac{y_{1i}}{p_{1i}} + \frac{1-\hat{P}}{\nu-m}\sum_{i=1}^{\nu-m}\frac{y_{2i}}{p_{2i}} \tag{2}$$

where $\hat{P} = (m-1)/(\nu-1)$ is an unbiased estimator of P.
Greco and Naddeo calculated $V(\tilde{T}_y)$ as following

$$V(\tilde{T}_y) = (W_1 - W_2)^2 V_\nu(\hat{P}) + \frac{\sigma_{1w}^2}{m}E_\nu(\hat{P}^2) + \frac{\sigma_{2w}^2}{m-1}E_\nu(\hat{P}(1-\hat{P})) \tag{3}$$

where $\sigma_{jw}^2 = \sum_{i=1}^{N_j}(W_{ji} - W_j)^2 p_{ji} \quad j = 1, 2$, and $W_{ji} = y_{ji}/pji$, $W_1 = T_1/P$, and $W_2 = T_2/(1-P)$.
An unbiased estimator of variance $V(\tilde{T}_y)$ is

$$\hat{V}(\tilde{T}_y) = (\bar{W}_1 - \bar{W}_2)^2\frac{\hat{P}(1-\hat{P})}{\nu-2} + \frac{S_{1w}^2}{m}\hat{P}_q + \frac{S_{2w}^2}{m-1}(\hat{P} - \frac{k-1}{m-2}\hat{P}_q))$$

where S_{jw}^2 is an unbiased sample variance of W_{ji}s in the jth group, and $\hat{P}_q = ((m-1)(m-2)/(\nu-1)(\nu-2))$ is an unbiased estimator of P^2. [4]
We now introduce a modified regression estimator for Greco and Naddeo's sampling design. To estimate B in inverse sampling with unequal selection probabilities we use estimators like (2). We use estimators $\tilde{T}_x/N$ and $\tilde{T}_y/N$ for $\bar{X}_N$ and $\bar{Y}_N$, respectively, and we estimate $\sum_{i=1}^{N} x_i y_i$ as following

$$\tilde{T}_{xy} = \frac{\hat{P}}{m}\sum_{i=1}^{m}\frac{x_{1i}y_{1i}}{p_{1i}} + \frac{1-\hat{P}}{\nu-m}\sum_{i=1}^{\nu-m}\frac{x_{2i}y_{2i}}{p_{2i}}.$$

If $\sum_{i=1}^{N} x_i^2$ is unknown, we can estimate it similarly as

$$\tilde{T}_{x^2} = \frac{\hat{P}}{m}\sum_{i=1}^{m}\frac{x_{1i}^2}{p_{1i}} + \frac{1-\hat{P}}{\nu-m}\sum_{i=1}^{\nu-m}\frac{x_{2i}^2}{p_{2i}}.$$

Then the modified regression estimator for inverse sampling with unequal selection probabilities is given by

$$\bar{y}_{UI.l} = \tilde{\mu}_y + b_{UI.l}(\bar{X}_N - \tilde{\mu}_x)$$

where

$$b_{UI.l} = \frac{\tilde{T}_{xy} - N\tilde{\mu}_x\tilde{\mu}_y}{\tilde{T}_{xy^2} - N\tilde{\mu}_x^2}.$$

To calculate variance $\bar{y}_{UI.l}$ we again assume B is known, then we have

$$\begin{aligned} V(\bar{y}_{UI.l}) &= V(\tilde{\mu}_y + B(\bar{X}_N - \tilde{\mu}_x)) = V(\tilde{\mu}_y - B\tilde{\mu}_x) \\ &= V(\frac{\hat{P}}{Nm}\sum_{i=1}^{m}\frac{y_{1i} - Bx_{1i}}{p_{1i}} + \frac{1-\hat{P}}{N(\nu-m)}\sum_{i=1}^{\nu-m}\frac{y_{2i} - Bx_{2i}}{p_{2i}}) \\ &= \frac{V(\tilde{T}_V)}{N^2} \end{aligned}$$

where $V = y - Bx$, and $V(\tilde{T}_V)$ can be calculated from (3). If we assume P is known, $V(\tilde{T}_V)$ can be simplified as following

$$V(\tilde{T}_V) = \frac{\sigma_{1w}^2}{m}P^2 + \frac{\sigma_{2w}^2}{m-1}P(1-P))$$

where $\sigma_{jw}^2 = \sum_{i=1}^{N_j}(W_{ji} - W_j)^2 p_{ji} \quad j = 1, 2$, and $W_{ji} = (y_{ji} - Bx_{ji})/pji$, $W_1 = (T_{1y} - BT_{1x})/P$, and $W_2 = (T_{2y} - BT_{2x})/(1-P)$.

3. Simulation study

Arsenic is a naturally occurring toxic element present in soil due to both natural and anthropogenic inputs. Arsenic-rich parent material and volcanic activity are the main natural, and mining and smelting process are the main anthropogenic source of AS. Kurdistan, a western province of Iran, is facing the problem of natural AS contamination. In the northeast of the province, there are some regions around Bijar, Qorveh and Takab cities where contaminated with AS (Karimi *et al.*, (2009) [6]).

A small number of studies have been conducted on AS concentration in groundwater and soils throughout this region (Mosaferi *et al.*, (2003) [7]; Modabberi and Moore (2004) [8]; and Karimi *et al.*, 2009).

The presence of AS in irrigation water or in soil at an elevated level could hamper normal growth of plants with the toxicity symptoms such as lower fruit and grain yield. Also AS found in edible parts of these plants cause deleterious effects for humans and other animals throughout eating of such fruits and vegetables. Determining AS concentration in plants is more complicated and more time-consuming experiment than determining AS concentration in water and soil (Karimi *et al.*, 2009).

Karimi *et al.* (2009) determined the AS concentration of some samples of groundwater, soil, and some plants in a contaminant region around Zarshuran mine of Takab. [6] We first partition the region in to 300 equal quadrates, let each quadrat is one sample unit. The contaminated region lies in 10 units. We then calculate

the mean of AS concentration of Karimi *et al.*'s samples in each contaminant unit for groundwater, soil, and plants separately. The results are shown in Table 1.

Table 1 Arsenic concentration in 10 contaminant units around Zarshuran mine of Takab in groundwater, soil, and plants.

z	x	y
130	224.84	8.52
200	299.76	3.66
500	407.70	1.39
1000	587.60	12.91
1860	757.47	20.50
2500	1127.30	21.62
3060	1541.85	23.77
3500	1487.10	33.64
4000	1667.00	37.13
16400	6105.55	155.76

Arsenic concentration in water is shown by z and it is measured by $\mu g/l$. Arsenic concentration in soil is shown by x and it is measured by mg/kg. Arsenic concentration in plant is shown by y and it is measured by mg/kg.

Standards of World Health Organization (WHO)(1993) says that AS concentration in safe (non-contaminated) water is below than 50 $\mu g/l$. [9] Based on such a standard we assume that $(z_i, x_i, y_i) = (25, 0, 0)$ for the remaining 290 safe units. In such a rare population we conduct a simulation study to determine the efficiency of estimators $\bar{y}_{GI.l}$ and $\bar{y}_{UI.l}$ in general inverse sampling design and inverse sampling design with unequal selection probabilities, respectively. For each estimator corresponding to $m = 2, 3, 4, 5$, we calculate

$$MSE(\bar{y}_\star) = \frac{1}{19000} \sum_{i=1}^{20000} (\bar{y}_\star - \bar{\bar{y}}_\star)^2 + (\bar{\bar{y}}_\star - \tau_y)^2$$

where $\bar{\bar{y}}_\star = \sum_{i=1}^{20000} \bar{y}_\star / 20000$ and $\star$ stands for the modified regression estimator in general inverse sampling, the conventional regression estimator with the same effective sample size in simple random sampling, the modified regression estimator in inverse sampling with unequal selection probabilities, and Greco and Naddeo's estimator that are respectively shown by $GI.l$, $SRS.l$, $UI.l$, and GN. Corresponding to each effective sample size $E(n) = \nu$, we calculate $V_{SRS}(\bar{y}) = (N - \nu)S^2/N\nu$ and efficiency of estimator $\bar{y}_\star$ which is given by

$$e(\bar{y}, \bar{y}_\star) = \frac{V_{SRS}(\bar{y})}{MSE(\bar{y}_\star)}.$$

Simulation results for general inverse sampling design and inverse sampling with unequal selection probabilities are summarized in Table 2 and Table 3, respectively. The summarized results in Table 2 show that the modified regression estimator in general inverse sampling $\bar{y}_{GI.l}$ is more efficient than conventional estimator in simple random sampling with an equal effective sample size. And $\bar{y}_{GI.l}$ is more efficient than regression estimator in simple random sampling $\bar{y}_{SRS.l}$ with an equal effective sample size. Also two last columns show that percent of samples that all their units are non-rare in general inverse sampling is too smaller than simple random sample one.

The simulation results in Table 3 shows that the modified regression estimator and Greco and Naddeo's estimator are more efficient than conventional estimator in simple random sampling with a same effective sample size, and the modified regression estimator is more efficient than Greco and Naddeo's estimator. In Table 3 the values of efficiencies are large because the variables z, x, and y are strongly strongly correlated, the correlation of each paired is more than 0.99.

An elementary comparison of Table 2 and Table 3 shows that when we have auxiliary variables that are correlated with the response variable using inverse sampling with unequal selection probabilities is more efficient than using general inverse sampling design because with a smaller sample size we gain very larger efficiency.

Table 2

Simulation of efficiencies in the AS containment population in Kurdistan province.

m	$e(\bar{y}_{GI.l})$	$e(\bar{y}_{SRS.l})$	$E(n)$	n_0	n_1	$p_{GI}(\underline{0})$	$p_{SRS}(\underline{0})$
2	25	6.6	58	40	100	1.7	11.3
3	24	8.7	74.8	40	100	1.5	5
4	21	10	88.1	40	100	1.8	2.8
4	31	13	110.6	80	150	0.09	1
5	20	11.6	95.7	40	100	1.4	2
5	28	17	128	90	150	0.1	0.4

The estimators $\bar{y}_{GI.l}$ and $\bar{y}_{SRS.l}$ are regression estimators in general inverse sampling and simple random sampling, respectively. And $p_{GI}(\underline{0})$ and $p_{SRS}(\underline{0})$ are the percent of samples that are selected just from non-rare subpopulation in general inverse sampling and simple random sampling, respectively.

Table 3

Simulation of efficiencies in the AS containment population in Kurdistan province.

m	$e(\bar{y}_{UI.l})$	$e(\bar{y}_{GN})$	$rb(\bar{y}_{UI.l})$	$E(n)$
2	721829	317339	-0.05	2.43
3	801387	315397	-0.03	3.64
4	808059	323608	-0.03	4.87
5	815476	322565	-0.02	6.1

The estimators $\bar{y}_{UI.l}$ and $\bar{y}_{GN}$ are regression estimator and Greco and Naddeo's estimator in inverse sampling with unequal selection probabilities. And $rb(\bar{y}_{UI.l})$ is relative biased of the modified regression estimator.

References

1. M. Moradi, M. M. Salehi and P. S. Levy. Inverse sampling design to avoid undefined estimator*Journal of Probability and Statistical Science* **5(2)**, 137 (2007).
2. J. B. S. Halden. On a method of estimating frequencies*Biometrika* **33**, 222 (1945).
3. M. M. Salehi and G. A. F. Seber. A general inverse sampling scheme and its application to adaptive cluster sampling*Australian and New Zealand Journal of Statistics* **43(3)**, 483 (2004).
4. L. Greco, and Naddeo. Inverse sampling with unequal selection probabilities*Communication in Statistics Theory and Methods.* **36(5)**, 1039 (2007).
5. W. G. Cochran, *Sampling techniques* (Wiley, New York, 1977).
6. N. Karimi, S. M. Ghaderian, H. Maroofi and H. Schot. Analysis of arsenic in soil and vegetation of a contaminated area in Zarshuran, Iran, two angiosperm arsenic hyperaccumulators.*International Journal of Phytoremediation.* In Press. (2009).
7. M. Mosaferi, M. Yunesion, A. Mesdaghinia, A. Naidu, S. Nasseri, A. Mahvi. Arsenic occurrence in drinking water of I.R. of Iran: the case of Kurdistan province. In: Ahmed MF, Ali MA, Adeel Z, editors. Fate of arsenic in the environment: proceedings of the BUET-UNU International Symposium, International Training Network Centre, Bangladesh University of Engineering and Technology, United Nations University, Tokyo, (2003). Dhaka: International Training Network Centre, Bangladesh University of Engineering and Technology, **1-6** (2003).
8. S. Modabberi and F. Moore. Environmental geochemistry of Zarshuran Au-As deposit, NW Iran*Environmental geology* **46 (6-7)**, 796 (2004).
9. WHO. Guidelines for drinking water quality: **2**. World Health Organization, Geneva. (1993).

Ridge Estimation In Singular Sur Models: New Approach

S. Najarian and M. Arashi

Faculty of Mathematics, Shahrood University of Technology

This paper proposes a procedure for developing unrestricted and restricted estimators of the seemingly unrelated regression parameters, when the explanatory variables are affected by multicollinearity. The new mechanism acts trough a singular regression setting. Some properties of new estimators are also derived.

Keywords: Multicollinearity, Restricted estimator, Seemingly unrelated regression, Singular linear model

1. Introduction

Multivariate regression requires the design matrix for each of p dependent variables to be the same in form. Zellner (1962) formulated Seemingly Unrelated Regression (SUR) models as p correlated regression equations. SUR models allow each of the p dependent variables to have a different design matrix with some of the predictor variables being the same. Of particular relevance to path analysis, SUR models allow for a variable to be both in the $\boldsymbol{Y}$ and $\boldsymbol{X}$ matrices. SUR models are a flexible analytic strategy and are underutilized in educational research. As a prelude to a SUR system, in the context of M simultaneous multiple regression equations, let

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \tag{1}$$

be the i^{th} equation of an M equation regression system with $\boldsymbol{y}_i$ a $T \times 1$ vector of observations on the i^{th} "dependent" variable, $\boldsymbol{X}_i$ a $T \times p_i$ matrix with rank l_i, of observations on p_i "independent" non-stochastic variables, $\boldsymbol{\beta}_i$ a $p_i \times 1$ vector of regression coefficients and $\boldsymbol{\epsilon}_i$, a$T \times 1$ vector of random error terms, each with mean zero. The system of which (1) is an equation may be written as:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & 0 & \dots & 0 \\ 0 & \boldsymbol{X}_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \boldsymbol{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_M \end{bmatrix}. \tag{2}$$

The full model can be rewritten as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3}$$

where for $n = MT$, $\boldsymbol{y}$ is a $(n \times 1)$ vector of responses and $\boldsymbol{X}$ is a $(n \times p)$ block-diagonal matrix, where $p = \sum_{i=1}^{M} p_i$. The $n \times 1$ disturbance vector in (2) and (3) is assumed to have the following variance-covariance matrix:

$$\begin{aligned}
\boldsymbol{\Sigma} &= Cov(\boldsymbol{\epsilon}) \\
&= \begin{bmatrix} \sigma_{11}\boldsymbol{I}_T & \sigma_{12}\boldsymbol{I}_T & \dots & \sigma_{1M}\boldsymbol{I}_T \\ \sigma_{21}\boldsymbol{I}_T & \sigma_{22}\boldsymbol{I}_T & \dots & \sigma_{2M}\boldsymbol{I}_T \\ \vdots & \vdots & & \vdots \\ \sigma_{M1}\boldsymbol{I}_T & \sigma_{M2}\boldsymbol{I}_T & \dots & \sigma_{MM}\boldsymbol{I}_T \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} & \sigma_{M2} & \dots & \sigma_{MM} \end{bmatrix} \otimes \boldsymbol{I}_T \\
&= \boldsymbol{\Sigma}_c \otimes \boldsymbol{I}_T
\end{aligned} \tag{4}$$

where $\boldsymbol{I}_T$ is a unit matrix of order $T \times T$ and $\sigma_{ii} = E(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it})$ for $t = 1, 2, \cdots, T$ and $i = 1, 2, \cdots, M$. In temporal cross-section regressions, t represents time and the model defined by (3) implies constant variances and covariances from period to period as well as the absence of any auto or serial correlation of the disturbances terms. Some interesting studies about SUR models include the couple of works done by Zellner (1962, 1963), Revankar (1976), Baltagi (1980), Srivastava and Giles (1978), Srivastvava and Maekawa (1995), Creel and Farell (1996), Moon and Perron (2004), Alkhamisi and Shukur (2008), Alkhamisi (2009).

One particularly important hypothesis in SUR models relates to test for aggregation bias, i.e.

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_M \quad \text{or} \quad \text{equivalently } H_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}, \tag{5}$$

where

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{I}_{p_1} & -\boldsymbol{I}_{p_2} & \boldsymbol{0}_{p_3} & \cdots \boldsymbol{0}_{p_{M-2}} & \boldsymbol{0}_{p_{M-1}} & \boldsymbol{0}_{p_M} \\ \boldsymbol{0}_{p_1} & \boldsymbol{I}_{p_2} & -\boldsymbol{I}_{p_3} & \cdots \boldsymbol{0}_{p_{M-2}} & \boldsymbol{0}_{p_{M-1}} & \boldsymbol{0}_{p_M} \\ \vdots & \vdots & \vdots & \dots \vdots & \vdots & \vdots \\ \boldsymbol{0}_{p_1} & \boldsymbol{0}_{p_2} & \boldsymbol{0}_{p_3} & \cdots \boldsymbol{I}_{p_{M-2}} & -\boldsymbol{I}_{p_{M-1}} & \boldsymbol{0}_{p_M} \\ \boldsymbol{0}_{p_1} & \boldsymbol{0}_{p_2} & \boldsymbol{0}_{p_3} & \cdots \boldsymbol{0}_{p_{M-2}} & \boldsymbol{I}_{p_{M-1}} & -\boldsymbol{I}_{p_M} \end{bmatrix}, \quad \boldsymbol{h} = \boldsymbol{0} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}$$

For the case of known matrix $\boldsymbol{\Sigma}_c$, it is well-known that under least squares (LS) theory, when nothing is known about the parameter space $\boldsymbol{\beta}$, the LS estimator

say, unrestricted estimator (UE) of $\boldsymbol{\beta}$ is given by

$$\dot{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}, \quad \boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}_c^{-1} \otimes \boldsymbol{I}_T. \tag{6}$$

Under occurrence of the linear restriction specified by (5), the restricted estimator (RE) of $\boldsymbol{\beta}$ has the form

$$\ddot{\boldsymbol{\beta}} = \dot{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{H}]^{-1}\boldsymbol{H}\dot{\boldsymbol{\beta}}. \tag{7}$$

Now suppose that in general, $\boldsymbol{\Sigma}$ is unknown. Then the estimate of $\boldsymbol{\Sigma}$ is given

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-q}(\boldsymbol{y} - \boldsymbol{X}\breve{\boldsymbol{\beta}})(\boldsymbol{y} - \boldsymbol{X}\breve{\boldsymbol{\beta}})', \qquad \breve{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{8}$$

From Zellner (1962), substituting (8) in equations (6) and (7), respectively gives the UE and RE of $\boldsymbol{\beta}$ as

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= (\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y} + O(T^{-1}),\\ \hat{\boldsymbol{\beta}} &= \tilde{\boldsymbol{\beta}} - (\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}'(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{H}]^{-1}\boldsymbol{H}\tilde{\boldsymbol{\beta}} + O(T^{-1}).\end{aligned} \tag{9}$$

In the forthcoming section, we define a new estimator based on the RE of $\boldsymbol{\beta}$ given by (9) under the presence of multicollinearity.

2. Singular SUR Model

For the consideration of a singular SUR system, under the assumptions of Section 1, based on the unified theory of LS proposed by Rao (1995), let

$$\boldsymbol{T} = \boldsymbol{\Sigma} + \boldsymbol{X}\boldsymbol{U}\boldsymbol{X}', \tag{10}$$

Where $\boldsymbol{U} \geq 0$, and $\text{rank}(T) = \text{rank}(\boldsymbol{\Sigma} \vdots \boldsymbol{X})$. Now, we define the following lagrangian criterion

$$F(\boldsymbol{\beta}, \lambda) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{T}^{-}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}'\boldsymbol{H}\boldsymbol{\beta},$$

Where $\boldsymbol{T}^{-}$ is the generalized inverse of $\boldsymbol{T}$.
By making use of the equation (9), it is easy to show that differentiating $F(\boldsymbol{\beta}, \lambda)$ by $\boldsymbol{\beta}$ and equating the derivatives to zero eventually yields the following RE of $\boldsymbol{\beta}$

$$\boldsymbol{\beta}_H^* = \boldsymbol{\beta}^* - (\boldsymbol{X}'\hat{\boldsymbol{T}}^{-}\boldsymbol{X})^{-1}\boldsymbol{H}'[\boldsymbol{H}(\boldsymbol{X}'\hat{\boldsymbol{T}}^{-}\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\boldsymbol{H}\boldsymbol{\beta}^* + O(T^{-1}), \tag{11}$$

where

$$\boldsymbol{\beta}^* = (\boldsymbol{X}'\hat{\boldsymbol{T}}^{-}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{T}}^{-}\boldsymbol{y} + O(T^{-1}), \quad \hat{\boldsymbol{T}} = \hat{\boldsymbol{\Sigma}} + \boldsymbol{X}\boldsymbol{U}\boldsymbol{X}'.$$

Now, we denote $\boldsymbol{S} = \boldsymbol{X}'\hat{\boldsymbol{T}}^{-}\boldsymbol{X}$ and $\boldsymbol{W} = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1}$, then $\boldsymbol{\beta}_H^*$ can be rewritten as

$$\boldsymbol{\beta}_H^* = \boldsymbol{W}\boldsymbol{X}'\hat{\boldsymbol{T}}^{-}\boldsymbol{y} + O(T^{-1}). \tag{12}$$

Multicollinearity is defined as the existence of nearly linear dependency among column vectors of the design matrix $\mathbf{X}^*$ in the linear model $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y}^*$ is an $n \times 1$ vector of observed responses, $\mathbf{X}^*$ is the observed matrix of independent variables of dimension $n \times p$, assumed to have full rank p, $\boldsymbol{\beta}$ is an unknown parameter, $\boldsymbol{\epsilon}$ is an error vector.

We know that the LS estimator of $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta} = \boldsymbol{T}^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{y}$, where $\boldsymbol{T} = \boldsymbol{X}^{*\prime}\boldsymbol{X}$. It is observed that the properties of the usual LS estimator of $\boldsymbol{\beta}$ depends heavily on the characteristics of the information matrix $\boldsymbol{T}$. If the matrix $\boldsymbol{T}$ moves from well-conditioned ones to ill-conditioned (near dependency among various columns of $\boldsymbol{T} = \boldsymbol{X}^{*\prime}\boldsymbol{X}^*$) or is collinear, then the LS estimator is sensitive to a number "errors", namely, there is an "explosion" of the sampling variance of the estimators. Moreover, some of the regression coefficients may be statistically insignificant with wrong sign and meaningful statistical inference becomes impossible for practitioners.

An adequate remedy for the effect of collinearity is to abandon the use of LS and use a biased estimation method known as ridge regression. To overcome collinearity under ridge regression, Hoerl and Kennard (1970) suggested the use of $\boldsymbol{T}_{(k)} = \boldsymbol{X}^{*\prime}\boldsymbol{X}^* + k\boldsymbol{I}_p$, $(0 \leq k \leq 1)$ rather than $\boldsymbol{T}$, in the estimation of $\boldsymbol{\beta}$, thereby developing the idea of ridge estimation of $\boldsymbol{\beta}$ given by $\boldsymbol{\beta}^*_{(k)} = \boldsymbol{T}^{(-1)}_{(k)}\boldsymbol{X}^{*\prime}\boldsymbol{y}$. The ridge regression approach has been studied by Hoerl and Kennard (1970), McDonald and Galarneau (1975), Lawless (1978), Gibbons (1981), Sarkar (1992), Saleh and Kibria (1993), Kibria and Saleh (2004) and Zhong and Yang (2007) to mention a few.

Now consider the existence of multicollinearity. In this regard, we modify equation (12) and propose a conditional ridge- type estimation

$$\boldsymbol{\beta}^*_H(k) = (k\boldsymbol{W} + \boldsymbol{I}_p)^{-1}\boldsymbol{\beta}^*_H + O(T^{-1}) \tag{13}$$

Where $k \geq 0$ is a constant (k is called conditional ridge parameter). If k chooses different values, we can get different estimator. Especially, $\boldsymbol{\beta}^*_H(0) = \boldsymbol{\beta}^*_H$ is the conditional best linear unbiased estimator of $\boldsymbol{\beta}$ when $k = 0$. In the forthcoming section, from Zhang and Yang (2007), we propose some properties of the estimator defined by (13) ignoring the terms which are of order $O(T^{-1})$, for the sake of simplicity. More developments with the presence of $O(T^{-1})$ and test of aggregating leave for further research.

3. Properties

In this section we keep on expressing an important superiority property of the proposed ridge estimator over its counterpart. Interested readers may refer to

Zhang and Yang (2007) for the rest.

Lemma 3.1. *(Wang, 1994) Suppose* $\boldsymbol{A}$ *is a real symmetric matrix, then* $\boldsymbol{A} \geq 0 \Leftrightarrow \forall \boldsymbol{P}, \boldsymbol{P}'\boldsymbol{A}\boldsymbol{P} \geq 0 \Leftrightarrow$ *each eigenvalue of* $\boldsymbol{A}$ *is non-negative.*

Lemma 3.2. $\boldsymbol{W} = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1} \geq 0$ *and rank*$(\boldsymbol{W}) = p - q$.

Proof: From the definition of $\boldsymbol{W}$, we have

$$
\begin{aligned}
&\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}' \\
&= [\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1}]\boldsymbol{S}[\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1}]' \\
&= \boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1} \\
&- \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{S}^{-1} \\
&+ \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1} \\
&= \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1} \\
&+ \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1} \\
&= \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{S}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{S}^{-1} \\
&= \boldsymbol{W}
\end{aligned} \tag{14}
$$

As $\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{T}^{-}\boldsymbol{X} \geq 0$ and from Lemma 3.1, it is obvious that $\boldsymbol{W} \geq 0$.
Now, we start to prove rank $(\boldsymbol{W}) = p - q$. Since $\boldsymbol{H}$ is a $q \times p$ restricted matrix with rank$(\boldsymbol{H}) = q$, then there exists $\boldsymbol{R}$, which is a $p \times p$ invertible matrix and satisfies $\boldsymbol{H} = (\boldsymbol{I}_{q\times q}, \boldsymbol{0}_{q\times p-q})\boldsymbol{R}$.
We divide $\boldsymbol{R}\boldsymbol{S}^{-1}\boldsymbol{R}'$ into:

$$
\boldsymbol{R}\boldsymbol{S}^{-1}\boldsymbol{R}' = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{21} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{bmatrix} \tag{15}
$$

Where $\boldsymbol{R}_{11}$ is a $q \times q$ matrix. So we have

$$
\begin{aligned}
\boldsymbol{RWR}' &= \boldsymbol{RS}^{-1}\boldsymbol{R}' - \boldsymbol{RS}^{-1}\boldsymbol{H}'(\boldsymbol{HS}^{-1}\boldsymbol{H})^{-1}\boldsymbol{HS}^{-1}\boldsymbol{R}' \\
&= \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{21} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{bmatrix} - \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{21} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{q\times q} \\ \boldsymbol{0} \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{I}_{q\times q} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{21} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{q\times q} \\ \boldsymbol{0} \end{bmatrix} \right)^{-1} \\
&\times \begin{bmatrix} \boldsymbol{I}_{q\times q} \\ \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{21} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{21} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{bmatrix} - \begin{bmatrix} \boldsymbol{R}_{11} \\ \boldsymbol{R}_{21} \end{bmatrix} \boldsymbol{R}_{11}^{-1}(\boldsymbol{R}_{11}, \boldsymbol{R}_{12}) \\
&= \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{21} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} \end{bmatrix} - \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{21} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_{22} - \boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12} \end{bmatrix}
\end{aligned}
$$

As $\boldsymbol{R}_{22}-\boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}$ is invertible, then $\text{rank}(\boldsymbol{W}) = \boldsymbol{R}(\boldsymbol{R}_{22}-\boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}) = p-q$.

Proposition 3.1. *$\boldsymbol{\beta}_H^*(k)$ is a shrinkage and biased estimator of $\boldsymbol{\beta}$.*

$$
\begin{aligned}
\|\boldsymbol{\beta}_H^*(k)\|^2 &= \|(k\boldsymbol{W}+\boldsymbol{I})^{-1}\boldsymbol{\beta}_H^*\|^2 \\
&= \|\boldsymbol{Q}(k\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}\boldsymbol{Q}'\boldsymbol{\beta}_H^*\|^2 \\
&= \|k\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}\boldsymbol{Q}'\boldsymbol{\beta}_H^*\|^2 \\
&< \|\boldsymbol{Q}'\boldsymbol{\beta}_H^*\|^2 \\
&= \|\boldsymbol{\beta}_H^*\|^2
\end{aligned}
$$

So $\boldsymbol{\beta}_H^*(k)$ is a shrinkage estimator.
As $E(\boldsymbol{\beta}_H^*(k)) = E((k\boldsymbol{W}+\boldsymbol{I})^{-1}\boldsymbol{\beta}_H^*) = (k\boldsymbol{W}+\boldsymbol{I})^{-1}\boldsymbol{\beta} \neq \boldsymbol{\beta}$, we can see $\boldsymbol{\beta}_H^*(k)$ is a biased estimator.

Proposition 3.2. *Under the meaning of Löwner partial ordering, the covariance of the ridge-type estimation $\boldsymbol{\beta}_H^*(k)$ is consistently superior over the covariance of the conditional estimator $\boldsymbol{\beta}_H^*$, that is*

$$Cov(\boldsymbol{\beta}_H^*(k)) \leq Cov(\boldsymbol{\beta}_H^*)$$

Proof: Since

$$
\begin{aligned}
Cov(\boldsymbol{\beta}_H^*) &= Cov(\boldsymbol{WX}'\boldsymbol{T}^-\boldsymbol{y}) \\
&= \sigma^2\boldsymbol{WX}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{XW}',
\end{aligned}
$$

and

$$\begin{aligned} Cov(\boldsymbol{\beta}_H^*(k)) &= Cov((k\boldsymbol{W}+\boldsymbol{I})^{-1}\boldsymbol{\beta}_H^*) \\ &= (k\boldsymbol{W}+\boldsymbol{I})^{-1}Cov(\boldsymbol{\beta}_H^*)(K\boldsymbol{W}+\boldsymbol{I})^{-1} \\ &= \sigma^2(k\boldsymbol{W}+\boldsymbol{I})^{-1}\boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}'(k\boldsymbol{W}+\boldsymbol{I})^{-1} \end{aligned}$$

We can get,

$$\begin{aligned} Cov(\boldsymbol{\beta}_H^*) - Cov(\boldsymbol{\beta}_H^*(k)) &= \sigma^2\boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}' \\ &\quad -\sigma^2(k\boldsymbol{W}+\boldsymbol{I})^{-1}\boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}'(k\boldsymbol{W}+\boldsymbol{I})^{-1} \\ &= \sigma^2(k\boldsymbol{W}+\boldsymbol{I})^{-1}[(k\boldsymbol{W}+\boldsymbol{I})\boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}'(k\boldsymbol{W}+\boldsymbol{I}) \\ &\quad -\boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}'](k\boldsymbol{W}+\boldsymbol{I})^{-1} \\ &= \sigma^2(k\boldsymbol{W}+\boldsymbol{I})^{-1}[k^2\boldsymbol{W}^2\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}^2 \\ &\quad +k\boldsymbol{W}^2\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}' + k\boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}^2](k\boldsymbol{W}+\boldsymbol{I})^{-1} \\ &= \sigma^2(k\boldsymbol{W}+\boldsymbol{I})^{-1}\boldsymbol{M}(k\boldsymbol{W}+\boldsymbol{I})^{-1} \end{aligned}$$

We note

$$\boldsymbol{M} = k^2\boldsymbol{W}^2\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}^2 + k\boldsymbol{W}^2\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}' + k\boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}^2$$

For $\boldsymbol{\Sigma} \geq \boldsymbol{0}, k \geq 0$ and from Lemma 3.1, we have

$$\begin{aligned} k^2\boldsymbol{W}^2\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}^2 \geq \boldsymbol{0}, \\ \boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}' \geq \boldsymbol{0} \end{aligned}$$

For $\boldsymbol{W} \geq \boldsymbol{0}$, and using Lemma 3.2 we have

$$\begin{aligned} k\boldsymbol{W}^2\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}' \geq \boldsymbol{0}, \\ k\boldsymbol{W}\boldsymbol{X}'\boldsymbol{T}^-\boldsymbol{\Sigma}\boldsymbol{T}^-\boldsymbol{X}\boldsymbol{W}^2 \geq \boldsymbol{0}. \end{aligned}$$

Thus we have $\boldsymbol{M} \geq \boldsymbol{0}$. So $Cov(\boldsymbol{\beta}_H^*(k)) \leq Cov(\boldsymbol{\beta}_H^*)$.

References

Alkhamisi, M. A. (2009). Simulation study of new estimators combining the SUR ridge regression and the restricted least squares methodologies, *Statist. Papers*, in press.

Alkhamisi, M. A. and Shukur, G. (2008). Developing Ridge Parameters for SUR Model, *Commun. Statist. - Theo. Meth.* ,**37**(4), 544-564.

Baltagi, B. (1980). On Seemingly Unrelated Regressions with Error Components, *Econometrica*, **48**, 1547-1552.

Creel, M., and Farell, M. (1996). SUR Estimation of Multiple Time-Series Models with Heteroskedasticity and Serial Correlation of Unknown Form, *Econ. Lett.*, **53**, 239-245.

Gibbons, D. G. (1981). A simulation study of some ridge estimators, *J. Amer. Statist. Assoc.*, **76**, 131-139.

Horel, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for non-orthogonal problems, *Technometrics*, **12**, 55-67.

Lawless, J. F. (1978). Ridge and related estimation procedure, *Comm. Statist. A*, **7**, 139-164.

McDonald, G. C. and Galarneau, D. I. (1975). A monte carlo evaluation of some ridge-type estimators, *J. Amer. Statist. Assoc.*, **70**, 407-416.

Moon, H. R. and Perron, P. (2004). Efficient Estimation of SUR Cointegration Regression Model and Testing for Purchasing Power Parity, *Econ. Rev.*, **23**, 293-323.

Rao, C. R. (1995). *Linear Models-Least Squares And Alternatives*, Springer-Verlag.

Revankar, N. S. (1976). Use of Restricted Residuals in SUR Systems: Some Finite Sample Results, *J. Amer. Statist. Assoc.*, **71**(353) 183-188.

Saleh, A. K. Md. E. and Kibria, B. M. G. (1993). Performances of some new preliminary test ridge regression estimators and their properties, *Comm. Statist. A*, **22**, 2747-2764.

Sarkar, N. (1992). A new estimator combining the ridge regression and the restricted least squares method of estimation, *Comm. Statist. A*, **21**, 1987-2000.

Srivastava, V. K. and Giles, D. E. A. (1987). *Seemingly Unrelated Regression Equations Models*, Marcel Dekker Inc., New York.

Srivastava, V. K. and Maekawa, K. (1995). Efficiency Properties of Feasible Generalized Least Squares Estimators in SURE Models under Non-normal Disturbances, *J. Econometrics*, **66**, 99-121.

Zhang, C. and Yang, H. (2007). The conditional ridge-type estimation in singular linear model with linear equality restrictions, *Statistics*, **41**(6), 485-494.

Zhong, Z. and Yang, H. (2007). Ridge estimation to the restricted linear model. *Comm. Statist. Theo.-Meth.*, **36**, 2099-2115.

Wang, S. (1994). *The inequality of Matrices*, Hefei: The Education of Anhui Press.

Zellner, A. (1963). Estimators for Seemingly Unrelated Regression Equations: Some Finite Sample Results, *J. Amer. Statist. Assoc.*, **58**, 977-992.

Zellner, A. (1962). An efficient method for estimating seemingly unrelated regressions and test for aggregation bias, *J. Amer. Statist. Assoc.*, **57**, 348-368.

A Review Of Statistical Methods In Split-Cluster Designs And Its Application In A Dental Clinical Trial Study

M. Namdari[1], H. Alavi Majd[1], R. Mofid[2] and Gh. Lotfi[2]

[1] *Department of Biostatistics, Shahid Beheshti University of Medical Sciences*
[2] *Faculty of Dentistry, Shahid Beheshti University of Medical Sciences*

In this paper we explain site data which are arising from clinical trials in periodontics and dermatology. In periodontal and dermatological trials it is relevant to frame a question in terms of sites than in terms of patients. Since host factors will influence all sites within a patient, it is obvious that sites within the same patient do not respond independently, so conventional statistical methods are not applicable and we should implement special statistical methods for their analysis. Split-cluster design is an extension of site or clustered data, which clusters such as multiple sites or organs in the same subject are assigned to different treatments. This design is popular in oral health research but most statisticians are not even aware of the existence of it. One of this design's advantages is that it removes a lot of inter-individual variability from the estimates of the treatment effect. In this article we explain and review the statistical methods for analyzing site data and split-cluster design when our response is binary, and also extend it to the case that we have baseline data in split-cluster design. The methodology is applied to an in vivo clinical trial for comparing the efficiency of a new treatment for gingival disease to a common treatment .

Keywords: Binary data, Dental clinical trials, GEE models , Interclass correlation, Site data, Split-Cluster design.

1. Introduction

Split-cluster designs are widely used by researchers in the health sciences when clusters such as multiple sites or organs on the same subject are assigned to different treatments (Donner et al (2004)). If all sites of an individual receive the same treatment, this is called the parallel-group design and it is the simplest and the most popular design in clinical trial methodology.

A frequently well known example of split cluster design is the "split-mouth" design which is adopted by periodontal researchers, in which the mouth is divided into two or more experimental units (segments) that are randomly assigned to different treatment groups (Donner and Zou (2007)). The investigators

may then be interested in comparing the overall proportion of sites (teeth or tooth surfaces) in each group that responds successfully to treatment. Other sciences in which this design can be applicable are experimental trials in dermatology (Bigby and Godenne (1986)) and animal studies (Weiss(2002)).

This design has the advantage of insulating treatment comparisons from inter-subject variability, and consequently have the potential to require many fewer subjects than a parallel arm trial having the same power(Donner and Zou (2007)and Emmanuel et al (2009)). Since the patient serves as his own control, it can increase statistical efficiency and on average fewer patients are needed.

When our outcome variable is continuous, statistical methods for split-cluster designs which are based on repeated measures analysis of variance, are well-known. There is a brief review on the appropriate statistical methods such as two way ANOVA and Paired t-test in the context of split-mouth designs by Hujoel and Moulton (1988). However methods for the analysis of binary outcome variables in split-cluster design are less developed.

In the next section of this paper we review characteristics of site or clustered data and the appropriate statistical methods for their analysis when the outcome variable is binary, in sections 3 we review different methods for analyzing split-cluster designs with binary response, in section 4 and 5 we review different methods for analyzing split cluster designs with binary response data that utilize baseline measurements, and at last we implement our methodology on a real clinical trial data for comparing the efficiency of a new treatment for gingival disease to a common treatment .

2. Site or Clustered data

At the beginning we should discuss some characteristics of site data. In some fields of research such as periodontal disease it is often more relevant to frame questions in terms of sites than in terms of patients, so we should perform the statistical analysis on a site basis. Sites measurements cannot be treated as independent within patients. It is obvious that patients factors influence all sites within a cluster so that sites within the same cluster will show association and are not independent, so statistical methods which assume independently distributed observations cannot be applied for analyzing site data. Here we determine a whole plot design in which all sites in a group of patients receive treatment "A" and all sites in the other group of patients receive treatment "B", and our response is in binary form.

If the sites within a patient could be regarded as independent, then the standard Pearson chi-square statistic (Fleiss (1981)), could be used to test the

equality of proportions in two treatment groups, but Pearson chi-square test is not a valid choice because of the clustering of sites within patients, and will lead to biased p-values. The magnitude of the bias will increase with the number of measurements per patient and with the size of the correlation among these measurements. Donner and Banting (1988) proposed an adjustment to the standard Pearson chi-square test that accounts for clustering effects in "whole-cluster" designs, in which all sites on a subject are assigned to either a treatment or a control condition

We assume it is of interest to compare 2 treatments on 2 groups of patients and we are interested about the prevalence of a specified characteristic over all sites of patients. Suppose the number of patients in the ith group is denoted by n_i, where $i = 1, 2$ denotes the experimental and control treatments, and patient j in group i contributes m_{ij} sites to the analysis, $j = 1, 2, ..., n_i$, where the response for each site can be classified as either a "success" or a "failure".

We assume that $\sum_i n_i$ subjects have been selected at random, where subject j,$j = 1, 2, ..., n_i$, contributes a total of m_j sites to the analysis. We also assume that treatment i, $i = 1, 2$, is randomly assign to all sites of the jth subject, where $i = 1$ denotes the experimental treatment and i = 2 denotes the control. So Y_{ijl} , $l = 1, 2, ..., m_{ij}$, $j = 1, 2, ..., n_i$, $i = 1, 2$, denote the observed binary outcome for lth site in jth subject who received treatment i, where $Y_{ijl} = 1$ denotes a "success" and $Y_{ijl} = 0$ denotes a "failure". $(Y_{ij} = \sum_l Y_{ijl})$ is the number of successes recorded on the m_{ij} sites assigned to treatment i on patient j . Donner and Banting (1988) assume that the Y_{ij} are independently distributed according to n independent beta-binomial distributions. The probability density for one of these distributions was defined by them as: $P(Y_{ij}) = \binom{m_{ij}}{Y_{ij}} \frac{\Gamma(\alpha_{ij}+\beta_{ij})\Gamma(Y_{ij}+\alpha_{ij})+\Gamma(m_{ij}-Y_{ij}+\beta_{ij})}{\Gamma(\alpha_{ij})\Gamma(\beta_{ij})\Gamma(m_{ij}+\alpha_{ij}+\beta_{ij})}$ where $\alpha_{ij} > 0$, $\beta_{ij} > 0$ are unknown parameters, and Γ represents the gamma function. This distribution is widely used to model dependent binary data. Donner and Banting (1988) interpreted it in this way that each individual in group i, i = 1,2 characterized by his own success rate.The observed number of successes Y_{ij} for individual j in this group is then assumed to follow a binomial distribution conditional on m_{ij} and the value of this rate. The further assumption that the individual success rates within a group vary according to a beta distribution yields a beta-binomial distribution with expected probability of success given by $P_i = \frac{\alpha_{ij}}{(1+\alpha_{ij}+\beta_{ij})}$.

Thus, the distribution (may be regarded as resulting from a mixture of binomial (within-subject) and extra-binomial (between-subject)variation. The existence of extra-binomial variation implies that the intraclass correlation ρ is positive, where ρ is the correlation coefficient between any two responses on the same

individual.

Here we will give some notations : We mentioned that Y_{ij} is the number of successes recorded on the m_{ij} sites assigned to treatment i on patient j, the corresponding success rate is then given by $\hat{P_{ij}} = \frac{Y_{ij}}{m_{ij}}$. We also denote the overall proportion of successes on treatment i by $\hat{P}_i = \frac{\sum_j Y_{ij}}{\sum_j m_{ij}} = \frac{Y_i}{M_i}$ and let $\hat{P} = \frac{\sum_i Y_i}{\sum_i M_i}$. We assume that there is a common correlation ρ among outcomes $(Y_{ijl}, Y_{ijl'})$ observed in the same cluster.

Observed outcomes from different clusters are assumed to be statistically independent.

In clinical trials our aim is to test the null hypothesis $H_0 : P_1 = P_2$ while accounting for the clustering effects, P_i is the underlying site-specific success rate for group i. For example, suppose we want to compare two treatments according to their ability to cure a specified infection. Then H_o states that the percentage of cured sites is the same for both treatments. Here we should understand that in some trials it would be more important to know whether treatment "A" was more successful in eliminating infections in a higher percentage of sites than treatment "B", than to know whether "A" totally eliminated infection in a higher percentage of patients.

If there is no extra-binomial variation, then $\rho = 0$ and we can test H_o by the standard Pearson chi-square statistic with (2-1) degrees of freedom.The Pearson chi-square statistic can be written as: $\chi^2_p = \frac{(\hat{P}_1 - \hat{P}_2)^2}{\hat{P}(1-\hat{P})(\frac{1}{M_1}+\frac{1}{M_2})}$ where $M_i = \sum_j m_{ij}$. If $\rho > 0$, χ^2 is no longer approximated by a chi-square distribution, but we can apply a modified form of it which has been introduced by Donner and Banting (1988).

They measure dependencies among observations in the same cluster by the intracluster correlation coefficient, and the "analysis of variance" estimator of ρ is $\hat{\rho} = \frac{(MSC - MSW)}{(MSC + (m_0 - 1)MSW)}$ where $m_0 = (M - \sum_i \bar{m}_{Ai})/(N-2)$, $N = \sum_i n_i$, $M = M_1 + M_2$, $MSC = \sum_i \sum_j m_{ij}((\hat{P}_{ij} - \hat{P}_i)/(N-2))$, $MSW = \sum_i \sum_j m_{ij}\hat{P}_{ij}(1 - \hat{P}_{ij})/(M-N)$, and $\bar{m}_{Ai} = \sum_j m^2_{ij}/M_i$ .They defined a clustering correction factor as $C_i = \sum_j m_{ij}[1 + (m_{ij} - 1)\hat{\rho}]/M_i$, $i = 1, 2$.

Donner and Banting (1988) adjusted chi-square statistic is, $\chi^2_A = (\hat{P}_1 - \hat{P}_2)^2/V_A$, where $V_A = \hat{P}(1-\hat{P})(C_1/M_1 + C_2/M_2)$. Under H_0, χ^2_A, approximately have chi-square distribution with one degree of freedom. We can proof easily that when $\hat{\rho} = 0$, χ^2_A reduces to the standard Pearson chi-square statistic χ^2_P, that we have introduced before.

3. Split-Cluster design

Split cluster design is an extension of clustered data or site data that we have discussed briefly in the previous section.We will explain a case in which each of k clusters is divided into exactly two segments, each segment can contain one or more sites (e.g., teeth), and the two segments in each cluster are randomly assigned to experimental or control treatment.

So we have k subjects that have been selected at random, and subject j, j = 1, 2,...,k, have a total of $m_{1j} + m_{2j}$ sites to the analysis. We also assume that treatment i, i = 1, 2, is randomly assigned to one of two segments of the jth subject, where i = 1 denotes the experimental treatment and i = 2 denotes the control. We define Y_{ijl} as the response for the lth site of the segment which receives treatment i, in the jth subject. In the previous part we mentioned that there is a correlation among outcomes which are in the same cluster, here in split cluster design there is a common correlation ρ among outcomes $(Y_{ijl}, Y_{ij\acute{l}})$ observed in the same segment, but we have another kind of correlation which is among outcomes $(Y_{1jl}, Y_{2j\acute{l}})$ observed in different segments within the same cluster and we showed it by ρ_{12}. We assume that outcomes from different clusters are statistically independent from each other. In clinical trials we want to test the null hypothesis $H_0 : P_1 = P_2$. For split cluster designs Donner et al (2004) on the basis of an extensive simulation study recommended to use a procedure which is based on a generalization of the standard Pearson chi-square statistic that takes into account both the clustering of sites within segments and the pair-matching of segments within subjects.

In their generalization form of χ^2_A they account for the correlation among matched pairs by adjusting the variance of $\hat{P}_1 - \hat{P}_2$ in the denominator of statistic. Donner et al (2004) has corrected the variance by $V_{GA} = \hat{P}(1-\hat{P})(C_1/M_1 + C_2/M_2 - 2\hat{\rho_{12}}\sum_j m_{1j}m_{2j}/(M_1M_2))$, where $\hat{\rho_{12}}$ may be obtained by computing the standard Pearson correlation over all possible pairs $(Y_{1jl}, Y_{2j\acute{l}})$ within a cluster, where each pair contains a single observation from each segment.

$$\hat{\rho_{12}} = \frac{\sum_{j=1}^{k}(Y_{1j} - m_{1j}\hat{P})(Y_{2j} - m_{2j}\hat{P})}{\sqrt{\sum_{j=1}^{k} m_{1j}(Y_{1j} - 2Y_{1j}\hat{P} + m_{1j}\hat{P}^2)\sum_{j=1}^{k} m_{2j}(Y_{2j} - 2Y_{2j}\hat{P} + m_{2j}\hat{P}^2)}} \quad (1)$$

Then the generalized version of the adjusted chi-square statistic is given by

$$\chi^2_{GA} = \frac{(\hat{P}_1 - \hat{P}_2)^2}{V_{GA}} \quad (2)$$

where ${\chi^2}_{GA}$ follows an approximate chi-square distribution with one degree of freedom under H_0.

Another method that is often reported in split-cluster designs for testing $H_0 : P_1 = P_2$ is the standard Paired t-test, this test can be applied to the differences $dj = \hat{P_{1j}} - \hat{P_{2j}}$. The test statistic is $t_P = \frac{\bar{d}}{S/\sqrt{k}}$ with $k-1$ degrees of freedom, which $\bar{d} = \sum_j d_j/k$ and$S^2 = \sum(d_j - \bar{d})^2/(k-1)$. We know that the normally distributed with equal variances assumptions of the d_js are violated, there is an alternative way and we can use nonparametric tests such as the Wilcoxon signed rank test but we would lose some power.

4. Statistical analysis of binary data in Split-Cluster design with baseline measurements

In many split-cluster designs, a baseline measurement may be obtained on each site within a segment. Since there is usually a high correlation between outcome measurement and its baseline measurement in site data, we can improve the efficiency of our analysis by accounting the baseline measurements in the analysis. We know that if we do not adjust an influential covariate in the analysis the statistical power will be reduced. However when our response variable is binary, methods for the statistical analysis of this kind of design are not developed. Donner and Zou (2007) presented several analytic approaches that can be implemented in such cases, they showed that the efficiency of statistical inferences can be improved by incorporating the baseline information.

In this part we introduce some methods that Donner and Zou (2007) was originally worked on it and can be used in cases with such kind of data . All notations are like the previous part, we introduce another notation similar to Y_{ijl} that accounts for baseline measurements; baseline measurements can be defined as X_{ijl}, with the superscript (0) used to denote success rates observed at baseline. For example, we define $P_{ij}^{(0)} = X_{ij}/m_{ij}$ and $P_i^{(0)} = \sum_j X_{ij}/\sum_j m_{ij}$, where $\hat{P}_i^{(0)}$ estimates the true proportion of successes for group i at baseline .

For testing the hypothesis that there is no differences between the two groups and taking into account the measurements at baseline Donner and Zou (2007) develop a test statistic based on change scores measured from baseline, they define a parameter for testing the null hypothesis that is: $\Delta = (P_1 - P_1^{(0)}) - (P_2 - P_2^{(0)})$ so the null hypothesis could be stated in this way: $H_0 : \Delta = 0$, and the test statistic would be;

$$\chi^2_{CH} = \frac{\hat{\Delta}^2}{V_{GAC}} \tag{3}$$

which its distribution is chi-square with one degree of freedom. V_{GAC} is the esti-

mated variance of $\hat{\Delta} = (\hat{P}_1 - \hat{P}_1^{(0)}) - (\hat{P}_2 - \hat{P}_2^{(0)})$,

$$V_{GAC} =$$
$$\hat{var}(\hat{P}_1 - \hat{P}_1^{(0)}) - 2\hat{\rho}_{12c}\sqrt{\hat{var}(\hat{P}_1 - \hat{P}_1^{(0)}).\hat{var}(\hat{P}_2 - \hat{P}_2^{(0)})} + \hat{var}(\hat{P}_2 - \hat{P}_2^{(0)}) \quad (4)$$

these variances can be estimated as

$$\hat{var}(\hat{P}_i - \hat{P}_1^{(0)}) = \hat{var}(\hat{P}_i) - 2\hat{cov}(\hat{P}_i, \hat{P}_1^{(0)}) + \hat{var}(\hat{P}_i^{(0)}) \quad (5)$$

where $\hat{var}(\hat{P}_i) = \frac{\sum_j (Y_{ij} - m_{ij}\hat{P}_i)^2}{(\sum_j m_{ij})^2}$, and $\hat{cov}(\hat{P}_i, \hat{P}_1^{(0)}) = \frac{\sum_j (Y_{ij} - m_{ij}\hat{P}_i)(X_{ij} - m_{ij}\hat{P}_i^{(0)})}{(\sum_j m_{ij})^2}$

In this design the correlation between two change scores observed in different segments within the same subject is denoted by $\rho_{12c} = corr(Y_{1jl} - X_{1jl}, Y_{2jl} - X_{2jl})$.The intrasubject correlation ρ_{12c} between a pair of change scores observed in different segments can be estimated in this way: first form all pairs of change scores for treatment ($d_{1jl} = Y_{1jl} - X_{1jl}$)and control ($d_{2jl} = Y_{2jl} - X_{2jl}$) within a subject and then compute the standard Pearson correlation over the set of all such pairs.

5. Generalized Estimating Equations approach

Another approach that can be useful for modeling data with baseline measurements in split cluster designs is the generalized estimating equations models. These models are not susceptible to problems that results from regression to the mean in the presence of baseline imbalances. GEE methods are initially introduced by Liang and Zeger (1986). GEE models are a good choice when we want to adjust our model for site-specific covariates and test their effects. GEE modeling relies on variance estimators constructed using between-cluster information, as a result, it requires a larger number of clusters as compared to χ^2_{CH} . In GEE modeling the binary follow-up measurement is used as the outcome variable and the baseline measurement and treatment group can be set as covariates.In the preceding model $T_{ijl} = 0, 1$ denote the control and experimental treatments and X_{ijl} is the baseline measurement for the lth site in the segment which received treatment i for the jth subject.

$$logit[Pr(Y_{ijl} = 1)] = \beta_0 + \beta_1 T_{ijl} + \beta_2 X_{ijl} \quad (6)$$

6. Application to a real data set

We implemented our statistical method in an in vivo study which was done by the Dental Research Center of Shahid Beheshti University. In this clinical trial five dogs with gingival disease which were met the inclusion criteria were participated in the study. The treatment procedure consists of isolating two teeth in each of the right

and left segments with cotton rolls and then have the treatment for the distal and mesial areas of the two teeth with either chitosan (control) or chitosan + fibroblast (experimental treatment). Determination of treatment condition (chitosan versus chitosan + fibroblast) for each segment was random and independent for each dog. The operator was unaware of the treatment that was applied. The measures associated with the width of gingiva were recorded at sites (distal and mesial) on each of the two selected teeth on the right and left segments, at baseline and after intervention. The left and right segments of the dogs jaw were randomly assigned to chitosan (control) or chitosan + fibroblast (treatment). Each treatment was applied to 4 sites(distal and mesial areas of the two teeth in each segment)so in this study $m = 8$ and $m_{1j} = m_{2j} = 4, k = 5$, if the measured width in in each site was less than 5 we recode it to 0 and if it was greater than 5 we recoded it to 1, codes were used for computing the statistics.

The sample estimates of P_1 and P_2, the proportion of patients with $width \geq 5$ in the control and experimental groups, respectively, are given by $P_1 = 0.3$ and $P_2 = 0.15$, with the intrasegment and intersegment correlation coefficients estimated as $\hat{\rho} = 0.23$ and $\hat{\rho}_{12} = -0.004$. SAS codes for computing the statistics were used.The values of the statistics for testing $H_0 : P_1 = P_2$ are given by $\chi^2_{GA} = 0.75$ which is not significant.

The sample estimates of $P_1^{(0)}$ and $P_2^{(0)}$, the proportion of patients with $width \geq 5$ in the control and experimental groups at baseline, respectively, are given by $P_1^{(0)} = 0$ and $P_2^{(0)} = 0.15$, with the intersegment correlation coefficients estimated as $\hat{\rho}_{12c} = 0$.The value of the statistic for testing $\Delta = 0$ is given by $\chi^2_{CH} = 5.29$ which is significant ($p = 0.025$), and it is more significant than that obtained when baseline measurements were ignored. We were not be able to estimate parameters in GEE model, because the generalized Hessian Matrix was not positive definite, we know that GEE modeling requires a large number of clusters to ensure its validity , in this example we have only five clusters and maybe that is the reason that we were not be able to estimate GEE parameters.

7. Discussion

A Monte Carlo simulation study that was implemented by Donner and Banting (1988) show that the adjusted chi-square test provides significance levels very close to nominal value over a wide range of parameter values. In the case of a constant $m_{ij} = m$, the clustering adjustment implied by this model is simply the division of the standard Pearson chi-square statistic by an estimate of the factor $1 + (m - l)\rho$, that in cluster sampling we known it as variance inflation factor .

A simulation study was conducted by Donner et all (2004)to compare the size and power of the test statistics that was described for split-cluster studies. For

the case of balanced cluster sizes, five sites in each cluster were assigned to either the intervention or control treatment. Their results show that the two procedures that we have mentioned for split cluster designs show empirical type I errors that are reasonably close to 0.05 under all parameter combinations considered in both the balanced and unbalanced cases. While imbalance in cluster size is seen to be accompanied by some loss in power for the both two procedures, but the effect on t_P was more severe. The statistics χ^2_{GA} show greater power than t_P at all parameter combinations.

When Baseline data are available, they can be used in order to increase the efficiency of the statistical analysis.Donner and Zou (2007) stated that χ^2_{CH} is a good choice for studies which enrolls 30 subjects or less, particularly if event rates are expected to be small. The simulation results by Donner and Zou (2007) show that GEE models tend to show significance levels that are somewhat unstable at n =20 and 30 when the baseline rates are equal to 0.1. However, GEE procedure provides acceptable type I error rates for all parameter values at n =50, so it requires a larger number of clusters to ensure its validity as compared to χ^2_{CH}, particularly when event rates are small. In the practical example that we used in this paper, we have only five clusters and maybe that is the reason that we were not be able to estimate GEE parameters,and it seems that it is more appropriate to interpret the value of χ^2_{CH} and χ^2_{CH} is the best choice.

8. Acknowledgement

The data for this work was supported by grants from the Dental Research Center of Shahid beheshti University, M.C, Tehran, Iran. We are grateful to them.

References

1. Bigby, M. Godenne, A. (1986). Continuing medical education: Understanding and evaluating clinical trials. *Journal of the American Academy of Dermatology,* 34, 555-593.
2. Donner, A. Banting, D.(1988). The statistical analysis of site-specific data in dental studies. *Journal of Dental Research,* 67:1392-1395.
3. Donner, A. Klar, N. Zou, G.(2004). Methods for the statistical analysis of binary data in split-cluster designs. *Biometrics,* 60,919-925.
4. Donner, A. Zou, G.(2007). Methods for the statistical analysis of binary data in split-mouth designs with baseline measurements. *Statistics in Medicine,* 26,3476-3486.
5. Fleiss, J.L. (1981). Statistical Methods for Rates and Proportions, 2th edition, John Wiley and Sons, New York .
6. Hujoel, P.P. Moulton, H.H. (1988).Evaluation of test statistics in split-mouth clinical trials. *Journal of Periodontology Research,* 23,378-380.

7. Lesaffre,E. Philstrom,B. Needleman,I. Worthington,H. (2009). The design and analysis of split-mouth studies: What statisticians and clinicians should know. *Statistics in Medicine,* 28,3470-3482.
8. Liang, KY. Zeger, SL.(1986). Longitudinal data analysis using generalized linear models. *Biometrika,* , 73,1322.
9. Weiss, R. A. (2002). Comparison of endovenous radiofrequency versus 810 nm diode laser occlusion of large veins in an animal model.*Dermatological Surgery,* 28, 56-61.

Using Wavelet Denoising for Solving Noisy-ICA Models

V. Nassiri and M. Aminghafari

Department of Mathematics, Vrije Universiteit Brussel (VUB), Brussels, Belgium.
Faculty Mathematics and Computer Sience, Amirkabir University of Technology

Independent Component Analysis (ICA) is a method for solving Blind Source Separation (BSS) problem. Since BSS is a real-world problem, its solution should be as realistic as possible. What always exists in real world is noise and therefore, it is very interesting to include noise in this model, we call this model noisy-ICA (NICA). In this report we try to use wavelets as natural denoising tools, in order to solve the NICA model. For this purpose we use the algorithm proposed by Aminghafari et al. [1] with a statistical approach.

Keywords: Independent Component Analysis; Noisy ICA; Wavelet Transform; Denoising; Fast ICA Algorithm.

1. Introduction

ICA is a multivariate statistical technique which has many applications in different fields which are extended from chemistry to music. However, most of the solutions of ICA do not consider noise in the model while it is inevitable in real-world problems, for a review of such solutions see e.g. Hyvärinen and Oja [6].

Wavelets are known as natural denoising tools, therefore, it is tempting to use them for solving the NICA model. There are some authors who used wavelets for this problem, e.g. Azzerboni et al. [2] and Tkacz [8]. Also we (Nassiri and Aminghafari [7]) proposed a method using wavelets for NICA in ISC9.

In this paper we try to apply a multivariate denoising method using wavelets and principal component analysis (PCA) introduced by Aminghafari et al. [1] in the NICA case. For this purpose we use a statistical approach using design of experiments (DOE) techniques for proposing an optimal algorithm, while it is not possible to find an optimal solution analytically, using statistical methods may give a useful answer to the question of optimal solution. The article is organized as follows. Section 2 will introduce the ICA model and its extension to the noisy case. In Section 3 we present the denoising algorithm and discuss how it can be used for solving NICA model. Our statistical approach for finding best solution of NICA using wavelet denoising will presented in Section 4. Section 5 is dedicated

to present the results of our study and finally in Section 6 the report is concluded.

2. ICA and NICA Models

In this section the ICA and NICA will be presented. Also the different kinds of noise and the performance of ICA-solving algorithms will be discussed.

2.1. *ICA and NICA*

Independent Component Analysis (ICA) is a recently developed solution for blind source separation (BSS) problem.

A simple description of BSS problem is a cocktail party. Consider two people are speaking simultaneously in a room. There are two microphones in different locations which record two signals. Obviously each of these signals is a mixture of the signals emitted by each person. In a linear approach, it's common to express this as linear equations as follows:

$$X_1 = a_{11}S_1 + a_{12}S_2$$

$$X_2 = a_{21}S_1 + a_{22}S_2$$

in matrix notation one can writes $X = AS$, where A is a square invertible mixing matrix[a]. Now the problem is finding S, just with knowing X. This is why it is called blind; since we have nothing more than an observed sample. If A is known, the problem is a simple linear system of equations, but in real world A is unknown.

In order to solve this problem by ICA, we assume that S_is $(i = 1, \ldots, k)$ are independent. This assumption is logical in many cases. ICA tries to find such an A which S_is are as independent as possible.

The general technique which is used to solve the ICA problem is considering a constraint which gets its extremum when the independence is true. Some famous constraints are achieved by maximizing non-Gaussianity. The motivation of this approach comes from central limit theorem (CLT). The most famous algorithm which uses the maximizing of non-Gaussianity is fastICA (Hyvärinen and Oja [5]).

One can easily extend the ICA model to the noisy case as follows,

$$X = AS + \epsilon \tag{1}$$

where ϵ is an additive noise with the following properties,

- ϵ is independent of the ICs
- The covariance matrix of ϵ is diagonal
- ϵ follows a Gaussian distribution

[a]This condition can be relaxed in many situations

- The covariance matrix of ϵ is known

in the rest of this atricle we try to exclude some of these assumptions which are not so realistic.

2.2. *Source noise and sensor noise*

Regarding the additive noise ϵ in Model (1), one may consider two kinds of noise. If you add the noise to X, i.e. $X = AS + \epsilon$ as in (1), it is called the sensor noise, but in some situations the source signals, S, may be noisy themselves, in this case the model is as follows,

$$X = A(S + \epsilon) = AS + A\epsilon = AS + \epsilon' \tag{2}$$

this model is known as source noise model. As one may see, the two models are the same at last, but one of the most interesting facts about source noise model is that the assumption said the covariance of ϵ is diagonal may no longer be true.

2.3. *Estimating the mixing matrix in NICA model*

The process of solving the ICA model has two main steps,

1. Estimating the mixing matrix A
2. Estimating the source signals S

now we concern on the first step. Gaussianity of the noise plays the main role in estimating A for NICA model. Because, as we denoted before, algorithms such as fastICA uses maximizing non-Gaussianity for estimating A. In fact, if a_i $(i = 1, \dots d)$ denotes the columns of A then in ordinary ICA, these methods try to find such an a_i which the non-Gaussianity of the vector $a_i^T X$ becomes maximum, where T denotes the transpose. For this purpose they use some indexes of non-Gaussianity such as Kurtosis or negentropy. Therefore, in the case of NICA we need to find such an a_i which non-Gaussianity of $a_i^T X + a_i^T \epsilon$ (or $a_i^T X + \epsilon$) becomes maximum. Fortunately, as Hyvärinen [4] showed, estimating A with maximizing non-Gaussianity using indexes such as kurtosis or negentropy does not be affected by the noise under certain conditions. In fact we can separate noisy signals, therefore, the last step is to extract original signals from noisy ones. In next section this matter will be discussed.

3. Extracting source signals using wavelet denoising

As in previous section denoted, algorithms such as fastICA can estimate the mixing matrix A regardless of noise. Now, if for the sake of simplicity we consider A as a square invertible matrix, then,

$$X = AS + A\epsilon \Rightarrow S = A^{-1}X - A^{-1}\epsilon = A^{-1}(X + \epsilon')$$

therefore, the noisy signals are separated, the only thing remains is denoising.

Wavelets are natural denoising tools, therefore, the straightforward solution is to apply discrete wavelet transform (DWT) on each column of $X+$ noise. But, this approach is useful when the covariance matrix of noise is diagonal, while as one may find out, it may not be true in NICA model, therefore, here we use an extension of denoising by wavelets proposed by Aminghafari et al. [1] which considers the correlations of noise. We call this algorithm here MWD (multivariate wavelet denoising), MWD has different versions, the one which is useful for our problem is as follows, here we denote the noisy source signals by S_n.

1. Perform the wavelet transform at level J of each column of X
2. Define $\hat{\Sigma_\epsilon}$ the estimator of noise covariance matrix as $\hat{\Sigma_\epsilon} = MCD(D_1)$ and then compute V such that $\hat{\Sigma_\epsilon} = V\Lambda V^T$ where $\Lambda = \text{diag}(\lambda_i)$. Apply to each detail after change of basis (namely $D_jV, 1 \leq j \leq J$), the p univariate thresholding strategies using the threshold $t_i = \sqrt{2\lambda_i \log(n)}$ for the i^{th} column of D_jV. Here p is the number of signals and n is the number of observations.
3. Reconstruct a denoised matrix $\check{S}$, from simplified detail and approximation matrices, by changing of basis using V^T and inverting the wavelet transform
4. Perform a final Principal Component Analysis (PCA) on the matrix $\check{S}$ obtained at step 4 and select $\check{p}$ PCs.

where $MCD(D_1)$ is applying a covariance estimating method introduced by Rousseeuw (1984) on D_1 which is the detail at level 1.

As it can be found by using this algorithm we exclude two assumptions on ϵ: Σ_ϵ is known and diagonal. However, there are some questions remains on how to use MWD and fastICA together, next section tries to answer them.

4. Some considerations on how to use MWD and fastICA together

A very common preprocessing step in using fastICa is making the signals white using some ordinary methods such as PCA or singular value decomposition (SVD), thus an estimation of the covariance matrix of X is needed. Therefore, in using MWD and fastICA for NICA model we have these options,

- First using fastICA for separating noisy signals and then denoising them using MWD, which may have two variations,
  - Using ordinary estimator of covariance matrix for whitening
  - Using $\text{E}(X'X) - MCD(D_1)$ for whitening
- First using MWD for denoising X and then using fastICA for separating denoised signals

choosing each of the above options, we need to choose at least three following items,

- Type of wavelet (DB, Sym, etc.)
- type of thresholding (hard or soft)
- How to perform thresholding (universal or level dependent)

Unfortunately, it is impossible or nearly-impossible to choose the best combination of those options analytically for a general problem. Here, we propose a practical way using statistical technique, design of experiments (DOE).

Here we want to find the best solution for NICA among those denoted above, by choosing the best wavelet and the best thresholding strategy. In the first step we need to choose the wavelet type, then we should choose between hard and soft thresholding and at last we should decide to threshold level dependent or universal. In fact, if we choose some levels for each factor, the effective factors in denoising have some levels which are nested in each other. Let choose DB4, Sym4 and Haar as the levels of wavelet type. Now for determining the effectiveness of each factor, a suitable design is a nested design. For collecting the information we need to use simulation, therefore, we need to determine dome distributions to simulate from. For sure the selected distribution will affect our results, therefore, the type of distribution is a nuisance factor and in order to exclude its effect from the results, we can consider it as a block factor, therefore, the final design is a nested block design. Let us to use Uniform$(0, 1)$, EXP(1), i.e. exponential distribution with parameter $\lambda = 1$, and $t(8)$, i.e. t-Student distribution with 8 degrees of freedom, as the levels of the nuisance factor. These three distributions can be regarded as representations of three type of distributions: symmetric, skew and uniform.

Remark 4.1. Since for two factors the type of wavelet and the type of distribution, we select just three among many choices, therefore, as it is usual in design and analysis of experiments we should consider them as factors with random effect, but because we choose them as representative ones not by chance, we consider them as fixed effect factors.

The last thing remains for constructing our model is the response variable. There are two important features for evaluating the goodness of NICA solution: good estimation of A and good estimation of source signals. Therefore, we also need two response variables for evaluating each of them.

For evaluating the goodness of estimation of A we use an index called Amari Distance (AmDi) introduced by Cichoki et al. (1996), here we use a normalized version of it which varied between 1 and 100 used by Barbedor [3]. Consider

$P = WA$ where W is the estimated demixing matrix, AmDi is as follows,

$$\text{AmDi} = \frac{100}{2d(d-1)} \left\{ \sum_{i=1}^{d} \left(\sum_{j=1}^{d} \frac{|P_{ij}|}{\max_k |P_{kj}|} - 1 \right) + \sum_{j=1}^{d} \left(\sum_{i=1}^{d} \frac{|P_{ij}|}{\max_k |P_{jk}|} - 1 \right) \right\} \tag{3}$$

where d is the number of components.

For evaluating the goodness of estimating of the source signals, Mean Square Error (MSE) is a natural choice, but for some restrictions in ICA model (for more information see e.g. Hyvärinen and Oja [6]), we propose to use MSE as follows,

$$\text{MSE} = \frac{1}{dn} \sum_{i=1}^{n} \left\{ \left(\sum_{j=1}^{d} S_j \right)_i - \left(\sum_{j=1}^{d} \hat{S}_j \right)_i \right\}^2 \tag{4}$$

The final model is as follows,

$$y_{ijklm} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(ij)} + \delta_{(ijk)l} + \epsilon_{(ijk)lm} \tag{5}$$

where μ is the overall effect of the treatments, α_i is the effect of chosen wavelet, $\beta_{j(i)}$ is the effect of choosing between soft and hard thresholding, $\gamma_{k(ij)}$ is the effect of choosing between level dependent and universal thresholding, $\delta_{(ijk)l}$ is the effect of block variable and $\epsilon_{(ijk)lm}$ is the error. Now, the only thing remains is to collect the data for our experiment and then analyze the results.

Data are simulated with 5 replications for each distribution and obviously not all of them are converged. We repeat the simulation for two different methods and also once for sensor noise and once for source noise. After collecting the data, the response variable for the case of MSE did not follow Gaussian distribution, therefore, ordinary analysis of variance was not possible, so we use some suitable non-parametric methods.

Based on the results, which are available if needed, the effective factors for both MSE and AmDi are as follows,

- Type of wavelet: **DB4**
- Type of thresholding: **hard**
- Strategy of thresholding: **level dependent**

Interestingly, the results obtained by our approach is in agreement with what we expect logically. Symmlet is mostly used in image processing, soft thresholding is used when the original signal is smooth while in the case of ICA it is a random sample of a non-Gaussian distribution and universal thresholding is a special case of level dependent thresholding.

Remark 4.2. In most of the denoising procedures, the original signal supposed to be smooth and the noise is random. The problem with ICA is that both the signal

and the noise are random, but fortunately the assumption of non-Gaussianity of the signal helps use for denoising by wavelets, since the threshold used before, can exclude a Gaussian noise and since the original signal is non-Gaussian the proposed method may have a good performance.

5. Results: the best solution for NICA

In this section we use simulation to find the best solution for NICA model. For this purpose, three different methods based on the combination of MWD and fastICA and also the method of whitening are considered.

1. MWD+fastICA
2. fastICA+MWD and using the ordinary estimation of variance of X for whitening
3. fastICA+MWD and using $\hat{\mathrm{E}}(X'X) - MCD(D_1)$ as the estimation of covariance of X for whitening.[b]

For performing simulation, we use 2600 replicates and apply each of the above methods on the data generated in each replication, then MSE, AmDi and the result of convergence of fastICA computed. For the distribution which data simulated from we select one of the three denoted distribution by chance (each with probability 1/3). The sample size in each replication was 1024 and also we have separate surveys on sensor and source noise. For wavelet denoising we use $J = 5$ which is usual for a signal with length 1024.

The results for sensor noise is presented in Table 1.

Method	AmDi	MSE	% convergence
MWD+ICA	41.09159	105549	93.96154
ICA+MWD†	21.0314	14789.38	98.26923
ICA+MWD‡	20.93285	15484.21	98.23077

† using the ordinary estimation of variance of X for whitening
‡ using $\hat{\mathrm{E}}(X'X) - MCD(D_1)$ for whitening

also the results for source noise is presented in Table 2.

According to the results superiority of ICA+MWD is obvious. Of course, the percentage of convergence of MWD+ICA is greater than ICA+MWD in source noise case. Another results which can be obtained is that there is no obvious preference between two method of whitening. Therefore, for choosing between

[b] using this method is called quasiwhitening

Method	AmDi	MSE	% convergence
MWD+ICA	40.44435	26152.75	93
ICA+MWD†	25.96535	11964.02	81.96154
ICA+MWD‡	25.67344	11844.02	81.34615

† using the ordinary estimation of variance of X for whitening
‡ using $\hat{\mathrm{E}}(X'X) - MCD(D_1)$ for whitening

them we need an index to use case to case for real-world data. We propose the signal to noise ratio (SNR) for this purpose. SNR is derived based on the fact that in practice one expects the variance of noise be larger than the variance of signal,

$$\mathrm{SNR} = \frac{\mathrm{Var}(\check{X})}{\mathrm{Var}(X - \check{X})} \tag{6}$$

it is obvious that a technique with less SNR is preferred.

Remark 5.1. Considering a note on why ICA+MWD is preferred to MWD+ICA may be interesting. One may know that the elements of X which are sums of some random variables have Gaussian distribution (according to the famous CLT). Now, when you add the Gaussian noise to them, the MWD cannot separate these two Gaussian vectors, because both of them are random samples. Therefore, performing the ICA first can solve this problem better.

6. Conclusions

In this article we used a statistical approach for determining the effective factors for using wavelets in NICA model. Also by simulation we find the best combination of fastICA and MWD algorithms. Based on the derived results we can propose the final algorithm with the following steps.

- Step1. Estimating the covariance of noise using $MCD(D_1)$ and call it $\hat{\Sigma}_\epsilon$
- Step2. Whitening data using $\mathrm{E}(X'X) - \hat{\Sigma}_\epsilon$ and call it $\tilde{X}$
- Step3. Performing fatsICA on X and denoising the derived matrix by MWD, then computing the mean of SNR for each of its columns and call it SNR1
- Step4. Performing fatsICA on $\tilde{X}$ without the whitening step and denoising the derived matrix by MWD, then computing the mean of SNR for each of its columns and call it SNR2
- Step5. Comparing SNR1 and SNR2, then choose the method with less SNR.

References

1. Aminghafari, M., Cheze, N., Poggi, J.M., (2006). Multivariate denoising using wavelets and principal component analysis. *Computational Statistics and Data Analysis*, **50**, 2381-2398.
2. Azzerboni, B., LA Forestn, F., Mammone, N. and Morabito, F. C., *A New Approach Based On Wavelet-ICA Algorithms For Fetal Electrocardiogram Extraction.* ESANN'2005, proceedings- European Symposium on Artificial Neural Networks, Brugge, Belgium, 2005, 27-29.
3. P. Barbedor, *Independent component analysis by wavelets*, Test, (2007) DOI 10.1007/s11749-007-0073-7.
4. A. Hyvärinen, *Gaussian moments for noisy independent component analysis*, IEEE signal processing letters, **6** (1999) 145-147.
5. A. Hyvärinen and E. Oja, *A fast fixed-point algorithm for indepedent component analysis*, Neural Computation, **13** (1997) 883-895.
6. A. Hyvärinen and E. Oja, *Indepedent component analysis: algorithms and applications*, Neural Computation, **13** (2000) 411-430.
7. Nassiri, V. and Aminghafari, M., *On suing wavelet thresholding for independent compoenent analysis.* Proceedings of ISC'9, 9^{th} Iranian Statics Conference, Isfahan University, Isfahan, Iran, 2009, 412-424.
8. Tkacz, E., Kostkal, P. and Mika, B., *An Application of Wavelet Transform (WT) and Independent Component Analysis (ICA) for Electrogastrographic (EGG) Signals Artifacts Detection.* Proceedings of World Congress on Medical Physics and Biomedical Engineering, Springer Verlag, 2007, 1026-1029.

An Overview of Composite Likelihood Methods: with emphasis on applications

A. Nematollahi

Department of Statistics, Shiraz University

The large number of parameters involved in the multivariate case makes the determination of maximum likelihood estimators much difficult and complicated, as it is not possible to compute the maximum likelihood estimator (MLE) of the parameters independently of variance-covariance matrix. In the case that the likelihood function is difficult to evaluate, certain methods based on modifications of the likelihood are used by several authors. In this work, we concentrate on the maximum pairwise likelihood estimator for the parameters in the three important cases: in symmetric normal model, in penalized linear models and finally in the multivariate AR(1) time series models. We observe that the loss of efficiency compared to maximum likelihood estimator is negligible, as tedious computations exist in formulating the later estimator.

1. Introduction

The large number of parameters involved in the multivariate case makes the determination of maximum likelihood estimators much difficult and complicated, as it is not possible to compute the maximum likelihood estimator (MLE) of the parameters independently of variance-covariance matrix. It amounts to solve tedious nonlinear equations involving variance-covariance matrix. In principle, maximum likelihood estimators can be computed with the aid of certain nonlinear optimization algorithms. In such procedures, it is important to begin the search with primary reasonable estimates. In the case that the likelihood function is difficult to evaluate, certain methods based on modifications of the likelihood are used by several authors.

Besag (1974, 1977) suggests pseudolikelihood for inference in spatial data. Pseudolikelihood is called composite likelihood by Lindsay (1988). Cox (1975) introduces the partial likelihood, and applies for fitting proportional hazard models. The pairwise likelihood, given by Cox and Reid (2004), takes the bivariate margins to produce the pseudolikelihood. For an excellent review on this topic with emphasis on some applications in genetic, longitudinal data, survival analysis and spatial statistics see Varin (2008).

In this paper, we concentrate on the maximum pairwise likelihood estimator

for the parameters in the three important cases: in symmetric normal model, in penalized linear models and finally in the multivariate AR(1) time series models. We observe that the loss of efficiency compared to maximum likelihood estimator is negligible, as tedious computations exist in formulating the later estimator.

2. Preliminaries

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)'$ be a m dimensional random vector with joint density $f(\mathbf{y}, \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbf{R}^m$ is unknown parameter. Under the usual regularity conditions and from n independent observations $\mathbf{Y}_1, ..., \mathbf{Y}_n$, the MLE of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}$, has an asymptotically multivariate normal distribution with mean $\boldsymbol{\theta}$ and variance-covariance the inverse of the expected Fisher information

$$\mathbf{J}(\boldsymbol{\theta}) = E\left[-\frac{\partial^2 \log f(\mathbf{Y}_1, ..., \mathbf{Y}_n, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right].$$

There are a number of situations where it is difficult to specify the full m-dimensional distribution in simple form but it maybe possible to specify some low-dimensional distributions. Here we consider the two-dimensional one, i.e., we concentrate for all $i, j = 1, ..., m$ the bivariate densities $f(y_i, y_j, \boldsymbol{\theta})$ for $i \neq j$. The pairwise likelihood (PL) from a single m-dimensional random vector $\mathbf{Y}$, is constructed by (Cox and Ried (2004))

$$L_{pl}(\mathbf{y}, \theta) = \prod_{i=1}^{m-1} \prod_{j=i+1}^{m} f(y_i, y_j, \boldsymbol{\theta}).$$

If the true parameter value $\boldsymbol{\theta}_0$ belongs to the interior of the compact parameter space then the PLE of $\boldsymbol{\theta}$, denoted by $\widetilde{\boldsymbol{\theta}}$, is the solution of the composite score function,

$$\mathbf{S}_2(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \frac{\partial \log(f(y_i, y_j, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = 0.$$

Since the composite score is a linear combination of valid likelihood score functions, then its unbiasedness follows under usual regularity conditions.

The information in the composite score equation provided by n observations is given by

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{J}_2(\boldsymbol{\theta}) \mathbf{K}_2^{-1}(\boldsymbol{\theta}) \mathbf{J}_2(\boldsymbol{\theta})$$

where

$$\mathbf{J}_2(\boldsymbol{\theta}) = E\left(-\frac{\partial \mathbf{S}(\boldsymbol{\theta}, \mathbf{y})}{\partial \boldsymbol{\theta}}\right),$$

and $\mathbf{K}_2(\boldsymbol{\theta})$, is the variance of the composite score matrix given by

$$\mathbf{K}_2(\boldsymbol{\theta}) = E[\mathbf{S}(\boldsymbol{\theta}, \mathbf{y})\mathbf{S}(\boldsymbol{\theta}, \mathbf{y})^T].$$

Quantity $\mathbf{G}(\boldsymbol{\theta})$ is known as Godambe information or sandwich information. Under some suitable regularity conditions, it can be shown that $\widetilde{\boldsymbol{\theta}}$ is consistent and asymptotically normal distributed with asymptotic mean $\boldsymbol{\theta}$, and variance matrix

$$\mathbf{G}^{-1}(\boldsymbol{\theta}) = \mathbf{J}_2^{-1}(\boldsymbol{\theta})\mathbf{K}_2(\boldsymbol{\theta})\mathbf{J}_2^{-1}(\boldsymbol{\theta}),$$

see Cox and Ried (2004) and Varin and Vidoni (2006). Note that, using pairwise likelihood, the information identity does not hold (since the model is misspecified), and hence the asymptotic variance matrix appears in the current form (Godambe or sandwich).

The asymptotic relative efficiency of $\widehat{\boldsymbol{\theta}}$ with respect to $\widetilde{\boldsymbol{\theta}}$ is therefore

$$\text{ARE}(\widehat{\boldsymbol{\theta}}|\widetilde{\boldsymbol{\theta}}) = \{\frac{\det(\mathbf{J}(\boldsymbol{\theta}))}{\det(\mathbf{G}(\boldsymbol{\theta}))}\}^{1/m},$$

and

$$\text{ARE}(\widehat{\theta}_r|\widetilde{\theta}_r) = \frac{(\mathbf{G}(\boldsymbol{\theta}))_{rr}}{(\mathbf{J}(\boldsymbol{\theta}))_{rr}}, r = 1, ..., m,$$

where $(\mathbf{C})_{rr}$ is the (r,r)th element of the inverse matrix $\mathbf{C}^{-1}$, see Davison, 2003.

3. Application 1: Symmetric normal model

Cox and Ried (2004) considered the following simple symmetric model; suppose

$$\mathbf{Y}_{q\times 1} \sim N_q\left(0, \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}\right),$$

so that every pair Y_s and Y_t has marginally a bivariate normal density with correlation ρ. The asymptotic variance of the usual MLE of ρ,, based on n such vectors, is

$$\frac{2}{nq(q-1)}\frac{\{1+(q-1)\rho\}^2(1-\rho)^2}{1+(q-1)\rho^2},$$

and the asymptotic variance based on the pairwise likelihood is

$$\frac{2}{nq(q-1)}\frac{1-\rho^2 c(q,\rho)}{(1+\rho^2)2},$$

where

$$c(q,\rho) = (1-\rho)^2(3\rho^2+1) + q\rho(-3\rho^3+8\rho^2-3\rho+2) + q^2\rho^2(1-\rho)^2.$$

Asymptotic relative efficiency of pairwise likelihood relative to full maximum likelihood, for $q = 3, 5, 8, 10$ (descending). This ratio is 1 for $q = 2$, as expected, and is also 1 if $\rho = 0$ or 1, for any value of q. Figure 1 illustrates the loss of information with increasing q.

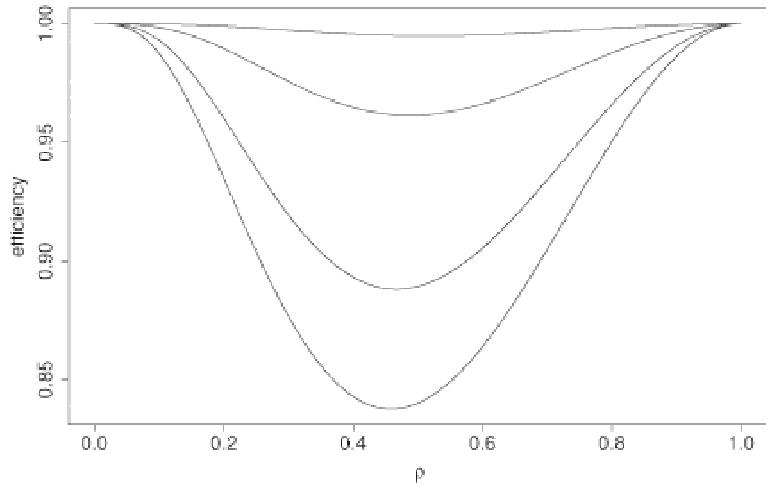

Fig. 1. Ratio of asymptotic variance of $\hat{\rho}$ to $\hat{\rho}_2$, as a function of ρ for fixed q. At $q = 2$ the ratio is identically 1. The lines shown are for $q = 3, 5, 8, 10$ (descending);

4. Application 2: Penalized composite likelihood

The following penalized model has been studied in an unpublished work by Nematollahi and Davison (2009). We have a response (observed) vector $\mathbf{y}$ of dimension n such that

$$\mathbf{y}|\mathbf{u} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \sim N_n\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma^2\mathbf{I}_n\right),$$

where

$\mathbf{X} = [x_{ij}]_{i=1,...,n;j=1,...,p}$: is a $n \times p$ (known) matrix of covariates,

$\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$: is an unknown corresponding $p \times 1$ parameter vector,

σ^2 : is an unknown nuisance parameter,

$\mathbf{Z} = [z_{ij}]_{i=1,...,n;j=1,...,q}$: is an $n \times q$ (known) matrix containing the elements of a basis, such as a collection of splines,

$\mathbf{u} = (u_1, ..., u_q)^T$: is an $q \times 1$ random coefficient and is treated as realization of a $N_q(\mathbf{0}, \boldsymbol{\Omega}_u)$.

(Indeed, $\mathbf{Zu}$ is a penalised component with corresponding coefficients $\mathbf{u}$).

Let $\mathbf{x}_i^T = (x_{i1}, ..., x_{ip})$ and $\mathbf{z}_i^T = (z_{i1}, ..., z_{iq}), i = 1, ..., n$.

We would like to estimate parameters $\boldsymbol{\beta}$, $\sigma^2$and Ω_u. First note that the marginal likelihood (the marginal distribution of $\mathbf{y}$) is given by

$$\begin{aligned}
f(\mathbf{y}; \mathbf{X}\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}_\mathbf{u}) &= \int f(\mathbf{y}|\mathbf{u}; \mathbf{X}\boldsymbol{\beta}, \sigma^2) f(\mathbf{u}; \boldsymbol{\Omega}_\mathbf{u}) d\mathbf{u} \\
&= \int (2\pi)^{-n/2} \left|\sigma^2 \mathbf{I_n}\right|^{-1/2} \\
&\quad \times \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Zu})^T (\sigma^2 \mathbf{I_n})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Zu})\right] \\
&\quad \times (2\pi)^{-q/2} |\Omega_u|^{-1/2} \left[-\frac{1}{2}\mathbf{u}^T \boldsymbol{\Omega}_\mathbf{u}^{-1} \mathbf{u}\right] d\mathbf{u} \\
&= ... \\
&= (2\pi)^{-n/2} \left|\sigma^2 \mathbf{I_n} + \mathbf{Z}\boldsymbol{\Omega}_\mathbf{u}\mathbf{Z}^T\right|^{-1/2} \\
&\quad \times \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\sigma^2 \mathbf{I_n} + \mathbf{Z}\boldsymbol{\Omega}_\mathbf{u}\mathbf{Z}^T)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \\
&= (2\pi)^{-n/2} |\boldsymbol{\Omega}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]
\end{aligned}$$

where $\boldsymbol{\Omega} = \sigma^2 \mathbf{I_n} + \mathbf{Z}\boldsymbol{\Omega}_\mathbf{u}\mathbf{Z}^T$, and so the marginal distribution of $\mathbf{y}$ (with $\mathbf{X}$ and $\mathbf{Z}$ are known) is

$$\mathbf{y} \sim \mathbf{N_n}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega}).$$

Note that $\boldsymbol{\Omega}$ depends on parameters $\psi = (\sigma^2, \boldsymbol{\Omega}_\mathbf{u})$ but not on $\boldsymbol{\beta}$.

4.1. *The likelihood and pairwise likelihood functions: Penalized composite likelihood*

For estimating the parameters, we consider two possibilities:

Case I: The parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}_{\mathbf{u}})$ are estimated by maximising the log marginal likelihood

$$l(\boldsymbol{\beta}, \boldsymbol{\Omega}) = l(\boldsymbol{\beta}, \psi) = l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}_{\mathbf{u}}) \propto -\frac{1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

We may find the maximum likelihood estimator of $\boldsymbol{\theta}$ which has under the usual regularity conditions an asymptotically normal distribution with mean $\boldsymbol{\theta}$ and variance the inverse of the expected Fisher information

$$\mathbf{J}(\boldsymbol{\theta}) = E\left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right].$$

Case II. We can also compute the corresponding pairwise marginal densities, which in this case just bivariate normal densities. Set $\mathbf{y}_{ij} := \binom{y_i}{y_j}$, we have

$$\mathbf{y}_{ij} \sim N_2\left(\mathbf{x}_{ij}\boldsymbol{\beta}, \boldsymbol{\Omega}_{ij}\right), i < j \in \{1, ..., n\},$$

where $\mathbf{x}_{ij} = \binom{\mathbf{x}_i^T}{\mathbf{x}_j^T}_{2\times p}$ and

$$\boldsymbol{\Omega}_{ij} = \sigma^2 \mathbf{I}_2 + \mathbf{z}_{ij} \boldsymbol{\Omega}_{u,ij} \mathbf{z}_{ij}^T.$$

where $\mathbf{z}_{ij} = \binom{\mathbf{z}_i^T}{\mathbf{z}_j^T}_{2\times q}, i < j = 1, ..., n$. Note that $\boldsymbol{\Omega}_{ij}$ depends on parameters $\psi_{ij} = (\sigma^2, \boldsymbol{\Omega}_{u,ij})$ but not on $\boldsymbol{\beta}$.Therefore the corresponding pairwise marginal density for $i < j \in \{1, ..., n\}$ is given by

$$f(y_i, y_i; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}_{u,ij}) = (2\pi)^{-1} |\boldsymbol{\Omega}_{ij}|^{-1/2} \exp -\frac{1}{2}\left\{(\mathbf{y}_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta})^T \boldsymbol{\Omega}_{ij}^{-1}(\mathbf{y}_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta})\right\}.$$

The log pairwise marginal likelihood is

$$\begin{aligned} l_2(\boldsymbol{\beta}, \boldsymbol{\Omega}_{ij}) = l_2(\boldsymbol{\beta}, \psi_{ij}) = l_2(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}_{u,ij}) &= \frac{1}{n-1}\sum_{i<j} \log f(y_i, y_i; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}_{u,ij}) \\ &= \frac{1}{n-1}\sum_{i<j}[-\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Omega}_{ij}| - \frac{1}{2}\left\{(\mathbf{y}_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta})^T \boldsymbol{\Omega}_{ij}^{-1}(\mathbf{y}_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta})\right\}]. \end{aligned}$$

The divisor $1/(n-1)$ is used so that the correct log likelihood is recovered when the observations are independent. Here $\boldsymbol{\Omega}_{ij}$ depends on parameters $\psi = (\sigma^2, \boldsymbol{\Omega}_{u,ij})$, so the ML equation $\boldsymbol{\Omega}_{ij}$ are obtained from equating to 0 the differentiation of l_2 with respect to $\psi_1 = \sigma^2$ and $\psi_2 = \boldsymbol{\Omega}_{u,ij}$.

We know the maximum pairwise likelihood estimator $\widetilde{\boldsymbol{\theta}}$ is consistent, and satisfied

$$\widetilde{\boldsymbol{\theta}} \sim N\left(\boldsymbol{\theta}, \overbrace{\mathbf{J}_2(\boldsymbol{\theta})^{-1}\mathbf{K}_2(\boldsymbol{\theta})\mathbf{J}_2(\boldsymbol{\theta})^{-1}}^{\text{sandwich variance}}\right),$$

as $n \to \infty$, where $\mathbf{J}_2(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(-\partial \mathbf{U}_2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(-\partial^2 l_2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T)$ is expected (pairwise) Fisher information matrix and $\mathbf{K}_2(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\mathbf{U}_2(\boldsymbol{\theta})\mathbf{U}_2(\boldsymbol{\theta})^T)$ is the variance of the score function.

4.2. *Main results in a simple case*

In the simplest case $\boldsymbol{\Omega}_{\mathbf{u}} = \sigma_u^2\mathbf{I}_q$,the information matrix for full models is

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{X}^T(\sigma^2\mathbf{I}_n + \sigma_u^2\mathbf{Z}\mathbf{Z}^T)^{-1}\mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}tr(\sigma^2\mathbf{I}_n + \sigma_u^2\mathbf{Z}\mathbf{Z}^T)^{-2} & \frac{1}{2}tr[(\sigma^2\mathbf{I}_n + \sigma_u^2\mathbf{Z}\mathbf{Z}^T)^{-2}\mathbf{Z}\mathbf{Z}^T] \\ \mathbf{0} & \frac{1}{2}tr[(\sigma^2\mathbf{I}_n + \sigma_u^2\mathbf{Z}\mathbf{Z}^T)^{-2}\mathbf{Z}\mathbf{Z}^T] & \frac{1}{2}tr[(\sigma^2\mathbf{I}_n + \sigma_u^2\mathbf{Z}\mathbf{Z}^T)^{-1}\mathbf{Z}\mathbf{Z}^T]^2 \end{bmatrix}$$

and for pairwise likelihood, we have

$$\mathbf{J}_2(\theta) = \begin{bmatrix} \frac{1}{n-1}\sum_{i<j}\mathbf{x}_{ij}^T(\sigma^2\mathbf{I}_2 + \sigma_u^2\mathbf{z}_{ij}\mathbf{z}_{ij}^T)^{-1}\mathbf{x}_{ij} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\frac{1}{n-1}\sum_{i<j}tr(\sigma^2\mathbf{I}_2 + \sigma_u^2\mathbf{z}_{ij}\mathbf{z}_{ij}^T)^{-2} & \frac{1}{2}\frac{1}{n-1}\sum_{i<j}tr[(\sigma^2\mathbf{I}_2 + \sigma_u^2\mathbf{z}_{ij}\mathbf{z}_{ij}^T)^{-2}\mathbf{z}_{ij}\mathbf{z}_{ij}^T] \\ \mathbf{0} & \frac{1}{2}\frac{1}{n-1}\sum_{i<j}tr[(\sigma^2\mathbf{I}_2 + \sigma_u^2\mathbf{z}_{ij}\mathbf{z}_{ij}^T)^{-2}\mathbf{z}_{ij}\mathbf{z}_{ij}^T] & \frac{1}{2}\frac{1}{n-1}\sum_{i<j}tr[(\sigma^2\mathbf{I}_2 + \sigma_u^2\mathbf{z}_{ij}\mathbf{z}_{ij}^T)^{-1}\mathbf{z}_{ij}\mathbf{z}_{ij}^T]^2 \end{bmatrix}$$

The variance of the score functionis

$$\mathbf{K}_2(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{(n-1)^2}\sum_{i<j}\sum_{k<l}\mathbf{x}_{ij}^T\boldsymbol{\Omega}_{ij}^{-1}\boldsymbol{\Phi}_{ijkl}\boldsymbol{\Omega}_{kl}^{-1}\mathbf{x}_{kl} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\sum_{i,j,k,l} tr[\boldsymbol{\Psi}_{ij}\boldsymbol{\Phi}_{ijkl}\boldsymbol{\Psi}_{kl}\boldsymbol{\Phi}_{ijkl}^T] & 2\sum_{i,j,k,l} tr[\boldsymbol{\Psi}_{ij}\boldsymbol{\Phi}_{ijkl}\boldsymbol{\Lambda}_{kl}\boldsymbol{\Phi}_{ijkl}^T] \\ \mathbf{0} & 2\sum_{i,j,k,l} tr[\boldsymbol{\Psi}_{ij}\boldsymbol{\Phi}_{ijkl}\boldsymbol{\Lambda}_{kl}\boldsymbol{\Phi}_{ijkl}^T] & 2\sum_{i,j,k,l} tr[\boldsymbol{\Lambda}_{ij}\boldsymbol{\Phi}_{ijkl}\boldsymbol{\Lambda}_{kl}\boldsymbol{\Phi}_{ijkl}^T] \end{bmatrix},$$

where

$$\boldsymbol{\Phi}_{ijkl} = \begin{pmatrix} \sigma^2\delta_{ik} + \sigma_u^2\mathbf{z}_i^T\mathbf{z}_k & \sigma^2\delta_{il} + \sigma_u^2\mathbf{z}_i^T\mathbf{z}_l \\ \sigma^2\delta_{jk} + \sigma_u^2\mathbf{z}_j^T\mathbf{z}_k & \sigma^2\delta_{jl} + \sigma_u^2\mathbf{z}_j^T\mathbf{z}_l \end{pmatrix},$$

$$\boldsymbol{\Psi}_{ij} = \begin{pmatrix} \Psi_{ij}^{11} & \Psi_{ij}^{12} \\ \Psi_{ij}^{12} & \Psi_{ij}^{22} \end{pmatrix} = \begin{cases} \mathbf{0} & i = j \\ \frac{1}{4(n-1)}\boldsymbol{\Omega}_{ij}^{-2} & i \neq j \end{cases},$$

$$\boldsymbol{\Lambda}_{ij} = \begin{pmatrix} \Lambda_{ij}^{11} & \Lambda_{ij}^{12} \\ \Lambda_{ij}^{12} & \Lambda_{ij}^{22} \end{pmatrix} = \begin{cases} \mathbf{0} & i = j \\ \frac{1}{4(n-1)}\boldsymbol{\Omega}_{ij}^{-1}\mathbf{z}_{ij}\mathbf{z}_{ij}^T\boldsymbol{\Omega}_{ij}^{-1} & i \neq j \end{cases},$$

and

$$\begin{aligned} \boldsymbol{\Omega}_{ij} &= \sigma^2\mathbf{I}_2 + \sigma_u^2\mathbf{z}_{ij}\mathbf{z}_{ij}^T = \sigma^2\begin{pmatrix} 1+\lambda z_{i.}^2 & \lambda\mathbf{z}_i^T\mathbf{z}_j \\ \lambda\mathbf{z}_i^T\mathbf{z}_j & 1+\lambda z_{j.}^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma^2 + \sigma_u^2\mathbf{z}_i^T\mathbf{z}_i & \sigma_u^2\mathbf{z}_i^T\mathbf{z}_j \\ \sigma_u^2\mathbf{z}_i^T\mathbf{z}_j & \sigma^2 + \sigma_u^2\mathbf{z}_j^T\mathbf{z}_j \end{pmatrix} \\ &= [\sigma^2\delta_{ij} + \sigma_u^2\mathbf{z}_i^T\mathbf{z}_j]_{i,j=1,2} \\ &= \sigma^2[\delta_{ij} + \lambda\mathbf{z}_i^T\mathbf{z}_j]_{i,j=1,2}. \end{aligned}$$

5. Application 3: Multivariate AR(1) models

5.1. *The multivariate AR(1) models*

Consider a m-variate time series $\left\{\mathbf{X}_t = (X_{t1}, X_{t2}, \ldots, X_{tm})^T, t \in \mathcal{Z}\right\}$, with $E\left(X_{ti}^2\right) < \infty,\ i = 1, 2, \ldots, m$ for all t, where $\mathcal{Z}$ stands for all integers. Let

$$\gamma_{ij}(t+h, t) = E\left[\left(X_{t+h,i} - \mu_{t+h,i}\right)\left(X_{t,j} - \mu_{t,j}\right)\right] \tag{1}$$

be the autocovariances, as the measure of dependency, not only between observations in the same series, but also between observation in the different series. Thus the second-order properties of the multivariate time series $\{\mathbf{X}_t\}$ are specified by the mean vectors,

$$\mu_t = E(\mathbf{X}_t) = E(X_{t1}, X_{t2}, \ldots, X_{tm})^T = (\mu_{t1}, \mu_{t2}, \ldots, \mu_{tm})^T \tag{2}$$

and autocovariance matrices,

$$\mathbf{\Gamma}(t+h, t) = E\left[\left(\mathbf{X}_{t+h} - \mu_{t+h}\right)\left(\mathbf{X}_t - \mu_t\right)^T\right] = \left[\gamma_{ij}(t+h, t)\right]_{i,j=1}^m \tag{3}$$

The m-variate time series $\{\mathbf{X}_t\}$ with mean and autocovariances (10) and (11) is said to be stationary if μ_t and $\Gamma(t+h, t), h = 0, \pm1, \pm2, \ldots$, are independent of t. For a stationary series we shall use the notation,

$$\mu = E(\mathbf{X}_t) = \left(\mu_1, \mu_2, \ldots, \mu_m\right)^T \tag{4}$$

and

$$\mathbf{\Gamma}(h) = E\left[\left(\mathbf{X}_{t+h} - \mu\right)\left(\mathbf{X}_t - \mu\right)^T\right] = \left[\gamma_{ij}(h)\right]_{i,j=1}^m \tag{5}$$

The autocorrelation matrix $\mathbf{R}(.)$is therefore

$$\mathbf{R}(h) = \left[\frac{\gamma_{ij}(h)}{\left(\gamma_{ii}(0)\,\gamma_{jj}(0)\right)^{1/2}}\right]_{i,j=1}^m = \left[\rho_{ij}(h)\right]_{i,j=1}^m \tag{6}$$

It is clear that $\mathbf{\Gamma}(h) = \mathbf{\Gamma}(-h)^T$.

Nematollahi and Kazemi (2009) consider the multivariate AR (1) models which provide a very useful class of models for describing the dynamics of a time series. A multivariate first order autoregressive models satisfies the following difference equation

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{Z}_t, \tag{7}$$

where $\mathbf{\Phi}$ is $m \times m$ matrix and $\mathbf{Z}_t$ is a sequence of independent multivariate normal with zero mean vector and covariance matrix Σ. Also we suppose that the process

is causal in the sense that all eigenvalues of $\mathbf{\Phi}$ are less than 1 in absolute value, or equivalently if all values of $z \in C$ satisfying

$$\det\left(\mathbf{I}_m - z\mathbf{\Phi}\right) = 0, \tag{8}$$

lie outside of unit circle.

5.2. *The pairwise likelihood inferences*

For a given set of m-dimensional observations $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ where $\mathbf{X}_t = (X_{t1}, X_{t2}, \ldots, X_{tm})^T$, $t = 1, ..., n$, we will gather all of them in a $mn \times 1$ vector $\mathbf{X}$ given by

$$\begin{aligned}\mathbf{X} &= \left(\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_n^T\right)^T \\ &= (X_{11}, X_{12}, \ldots, X_{1m,} X_{21}, X_{22}, \ldots, X_{2m,} \ldots, X_{n1}, X_{n2}, \ldots, X_{nm})^T.\end{aligned}$$

The pairwise likelihood for $\mathbf{X}$ can be written as

$$\begin{aligned}L_{pl}\left(\mathbf{x}, \boldsymbol{\theta}\right) = &\left[\prod_{a=1}^{n}\prod_{b=1}^{m-1}\prod_{c=b+1}^{m} f_{(X_{ab}, X_{ac})}\left(x_{ab}, x_{ac}, \boldsymbol{\theta}\right)\right] \\ &\times \left[\prod_{i=1}^{n-1}\prod_{j=i+1}^{n}\prod_{b=1}^{m}\prod_{c=1}^{m} f_{(X_{ib}, X_{jc})}\left(x_{ib}, x_{jc}, \boldsymbol{\theta}\right)\right]\end{aligned} \tag{9}$$

Note that in order to use all of the information in the time series, we have to write $\mathbf{X}$ and $L_{pl}\left(\mathbf{x}, \boldsymbol{\theta}\right)$ as above, since in the multivariate time series the vector of observations at different times are dependent and it is also dependent with other observations. In (9), the first term is the likelihood function of the observations within the same groups and the second term is the counterpart between different groups.

For the stationary causal multivariate AR (1) process, the exact likelihood is

$$\begin{aligned}L\left(\mathbf{X}, \mathbf{\Phi}, \mathbf{\Sigma}\right) &= \left(2\pi\right)^{-nm/2}\left(\det\left(\mathbf{\Gamma}\left(0\right)\right)\right)^{-1/2}\left(\det\left(\mathbf{\Sigma}\right)\right)^{-(n-1)/2} \\ &\quad \times \exp\left[\frac{-1}{2}\left(\mathbf{X}_1^T\mathbf{\Gamma}\left(0\right)^{-1}\mathbf{X}_1 + \sum_{j=1}^{n}\mathbf{Z}_t^T\mathbf{\Sigma}^{-1}\mathbf{Z}_t\right)\right] \\ &= f_1\left(\mathbf{X}_1\right) L_*\left(\mathbf{X}_2, \mathbf{X}_3, \ldots, \mathbf{X}_n, \mathbf{\Phi}, \mathbf{\Sigma}\right),\end{aligned} \tag{10}$$

where
$L_*\left(\mathbf{X}_2, \mathbf{X}_3, \ldots, \mathbf{X}_n, \mathbf{\Phi}, \mathbf{\Sigma}\right)$ represent the conditional p.d.f. of $\left(\mathbf{X}_2, \mathbf{X}_3, \ldots, \mathbf{X}_n\right)$ given $\mathbf{X}_1$, and is the conditional likelihood.

Also note that, with $\boldsymbol{\phi} = vec\left(\mathbf{\Phi}\right)$ then $\widehat{\boldsymbol{\phi}}$ (the MLE of $\boldsymbol{\phi}$) is approximately normal distributed with mean $\boldsymbol{\phi}$ and variance covariance matrix $1/n(\mathbf{\Gamma}\left(0\right)^{-1} \otimes \mathbf{\Sigma})$, see Reinsel(1997) for more details.

By the above discussion, for a nm-dimensional Gaussian vector $\mathbf{X}$, we can say

$$\begin{bmatrix} X_{ab} \\ X_{ac} \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \gamma_{bb}\,(0)\ \gamma_{bc}\,(0) \\ \gamma_{bc}\,(0)\ \gamma_{cc}\,(0) \end{bmatrix}\right)$$
$$\begin{bmatrix} X_{ib} \\ X_{jc} \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \gamma_{bb}\,(i-j)\ \gamma_{bc}\,(i-j) \\ \gamma_{bc}\,(i-j)\ \gamma_{cc}\,(i-j) \end{bmatrix}\right). \tag{11}$$

Then we can use (11) for computing the pairwise likelihood function $L_{pl}\,(\mathbf{x}, \boldsymbol{\theta})$.

For notation simplicity take

$$M_{bc}(h) = \begin{bmatrix} \gamma_{bb}(h)\ \gamma_{bc}(h) \\ \gamma_{bc}(h)\ \gamma_{cc}(h) \end{bmatrix}$$

so

$$\begin{bmatrix} X_{ib} \\ X_{jc} \end{bmatrix} \sim N\ \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}\ ,\ M_{bc}(i-j)\right].$$

Then we can write $L_{pl}(\boldsymbol{\theta}, X)$ in the following form

$$\begin{aligned} l_2(\boldsymbol{\theta}) &:= \log(L_{pl}(X, \boldsymbol{\theta})) \\ &\propto \sum_{a=1}^{n}\sum_{b=1}^{m-1}\sum_{c=b+1}^{m} [(-1/2)\log|M_{bc}(0)| \\ &\quad -(1/2)\left\{X_{bc}^T(a)M_{bc}^{-1}(0)X_{bc}(a)\right\}] \\ &\quad +\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{b=1}^{m}\sum_{c=1}^{m}[(-1/2)\log|M_{bc}(i-j)| \\ &\quad -(1/2)\left\{X_{bc}^T(i,j)M_{bc}^{-1}(i-j)X_{bc}(i,j)\right\}], \end{aligned} \tag{12}$$

where $\boldsymbol{\theta} = (\boldsymbol{\Phi}, \Sigma)$ is the unknown parameter of model, $X_{bc}(a) = \begin{pmatrix} X_{ab} \\ X_{ac} \end{pmatrix}$ and $X_{bc}(i,j) = \begin{pmatrix} X_{ib} \\ X_{jc} \end{pmatrix}$.

Now let $\boldsymbol{\Phi} = [\phi_{ij}]_{i,j=1}^{m}$ and $\boldsymbol{\Sigma} = [\sigma_{ij}]_{i,j=1}^{m}$ and take $w_1 = \phi_{11},\ w_2 = \phi_{21},\ ...,\ w_m = \phi_{m1},\ ...,\ w_{mm+(m(m+1)/2)} = \sigma_{mm}$. Then, the composite (pairwise) score function is

$$\begin{aligned} \mathbf{S}_2(\boldsymbol{\theta}) &= \mathbf{S}_2(w_1,\ w_2\ ...,\ w_m,\ ...,\ w_{mm+(m(m+1)/2)}) \\ &= \left[\frac{\partial l_2(\boldsymbol{\theta})}{\partial w_1}, \frac{\partial l_2(\boldsymbol{\theta})}{\partial w_2}, ..., \frac{\partial l_2(\boldsymbol{\theta})}{\partial w_m}, ..., \frac{\partial l_2(\boldsymbol{\theta})}{\partial w_{mm+(m(m+1)/2)}}\right]_{1\times(mm+(m(m+1)/2))} \end{aligned} \tag{13}$$

and

$$\mathbf{J}_2(\boldsymbol{\theta}) = \left[E\left(\left[\frac{\partial^2 l_2(\boldsymbol{\theta})}{\partial w_k \partial w_l} \right] \right) \right]_{k,l=1}^{mm+(m(m+1)/2)}, \tag{14}$$

and

$$\mathbf{K}_2(\boldsymbol{\theta}) = \left[E\left(\left[\frac{\partial l_2(\boldsymbol{\theta})}{\partial w_k} \right] \left[\frac{\partial l_2(\boldsymbol{\theta})}{\partial w_l} \right] \right) \right]_{k,l=1}^{mm+(m(m+1)/2)}. \tag{15}$$

First note that

$$\begin{aligned} \frac{\partial l_2(\boldsymbol{\theta})}{\partial w_k} = & \sum_{a=1}^{n} \sum_{1\le b<c\le m} [(-1/2)\left\{ \frac{\partial \log |M_{bc}(0)|}{\partial w_k} \right\} \\ & -(1/2)\left\{ X_{bc}^T(a) \left\{ \frac{\partial M_{bc}^{-1}(0)}{\partial w_k} \right\} X_{bc}(a) \right\}] \\ & + \sum_{1\le i<j\le n}^{n-1} \sum_{b,c=1}^{m} [(-1/2)\left\{ \frac{\partial \log |M_{bc}(i-j)|}{\partial w_k} \right\} \\ & -(1/2)\left\{ X_{bc}^T(i,j) \left\{ \frac{\partial M_{bc}^{-1}(i-j)}{\partial w_k} \right\} X_{bc}(i,j) \right\}], \end{aligned} \tag{16}$$

and then we have,

$$\begin{aligned} \frac{\partial^2 l_2(\boldsymbol{\theta})}{\partial w_k \partial w_l} = & \sum_{a=1}^{n} \sum_{1\le b<c\le m} [(-1/2)\left\{ \frac{\partial^2 \log |M_{bc}(0)|}{\partial w_k \partial w_l} \right\} \\ & -(1/2)\left\{ X_{bc}^T(a) \left\{ \frac{\partial^2 M_{bc}^{-1}(0)}{\partial w_k \partial w_l} \right\} X_{bc}(a) \right\}] \\ & + \sum_{1\le i<j\le n}^{n-1} \sum_{b,c=1}^{m} [(-1/2)\left\{ \frac{\partial^2 \log |M_{bc}(i-j)|}{\partial w_k \partial w_l} \right\} \\ & -(1/2)\left\{ X_{bc}^T(i,j) \left\{ \frac{\partial^2 M_{bc}^{-1}(i-j)}{\partial w_k \partial w_l} \right\} X_{bc}(i,j) \right\}]. \end{aligned} \tag{17}$$

Now we can state the following theorems that play the main roles in computing the asymptotic variance-covariance matrix of the PLE. The proof of these theorems are left to Nematollahi and Kazemi (2009), because of the tedious computations.

Theorem 5.1. A typical (k,l)th element of the matrix $\mathbf{J}_2(\theta)$ has the form

$$\begin{aligned} E\left(\frac{\partial^2 l_2(\boldsymbol{\theta})}{\partial w_k \partial w_l}\right) = \sum_{a=1}^{n} \sum_{1\le b<c\le m} [(1/2)\left\{\frac{\partial^2 \log|M_{bc}(0)|}{\partial w_k \partial w_l}\right\} \\ +(1/2)tr\left(\left\{\frac{\partial^2 M_{bc}^{-1}(0)}{\partial w_k \partial w_l}\right\} M_{bc}(0)\right)] \\ + \sum_{1\le i<j\le n}^{n-1} \sum_{b,c=1}^{m} [(1/2)\left\{\frac{\partial^2 \log|M_{bc}(i-j)|}{\partial w_k \partial w_l}\right\} \\ +(1/2)tr\left(\left\{\frac{\partial^2 M_{bc}^{-1}(i-j)}{\partial w_k \partial w_l}\right\} M_{bc}(i-j)\right)] \, . \end{aligned}$$

Theorem 5.2. A typical (k,l)th element of the variance of the composite score matrix $\mathbf{K}_2(\theta)$, has the form

$$\begin{aligned} & E\left[(\partial l_2(\Theta)/\partial w_k)\,(\partial l_2(\Theta)/\partial w_l)\right] = \sum_{a=1}^{n} \sum_{1\le b<c\le m} \sum_{a'=1}^{n} \sum_{1\le b'<c'\le m} \\ & [(1/2)tr\left[N_{bc,k}(0) M_{bb'cc'}(a-a') N_{b'c',l}(0) M_{bb'cc'}^{T}(a-a')\right] \\ & + \sum_{1\le i<j\le n} \sum_{b,c=1}^{m} \sum_{a'=1}^{n} \sum_{1\le b'<c'\le m} \\ & [(1/2)tr\left[N_{bc,k}(0) M_{bb'cc'}(a-i', a-j') N_{b'c',l}(i'-j') M_{bb'cc'}^{T}(a-i', a-j')\right] \\ & \sum_{1\le i<j\le n} \sum_{b,c=1}^{m} \sum_{a'=1}^{n} \sum_{1\le b'<c'\le m} \\ & (1/2)tr\left[N_{bc,k}(i-j) M_{bb'cc'}(i-a',\ j-a') N_{b'c',l}(0) M_{bb'cc'}^{T}(i-a',\ j-a')\right] \\ & + \sum_{1\le i<j\le n} \sum_{b,c=1}^{m} \sum_{1\le i'<j'\le n} \sum_{b',c'=1}^{m} \\ & (1/2)tr[N_{bc,k}(i-j) M_{bb'cc'}(i-i', i-j', j-i', j-j') \\ & N_{b'c',l}(i'-j') M_{bb'cc'}^{T}(i-i', i-j', j-i', j-j')], \end{aligned}$$

where $N_{bc,k}(h) = \partial M_{bc}^{-1}(h)/\partial w_k$, $M_{bb'cc'}(h) = \begin{bmatrix} \gamma_{bb'}(h) & \gamma_{bc'}(h) \\ \gamma_{cb'}(h) & \gamma_{cc'}(h) \end{bmatrix}$,

$M_{bb'cc'}(h,h') = \begin{bmatrix} \gamma_{bb'}(h) & \gamma_{bc'}(h') \\ \gamma_{cb'}(h) & \gamma_{cc'}(h') \end{bmatrix}$

and $M_{bb'cc'}(h_1,h_2,h_3,h_4) = \begin{bmatrix} \gamma_{bb'}(h_1) & \gamma_{bc'}(h_2) \\ \gamma_{cb'}(h_3) & \gamma_{cc'}(h_4) \end{bmatrix}$.

Finally by using Theorems 5.1, 5.2 and, the asymptotic variance-covarianve matrix can be derived and then we can compute the asymptotic relative efficiency of $\widehat{\boldsymbol{\theta}}$ with respect to $\widetilde{\boldsymbol{\theta}}$. For more details, refer to Nematollahi and Kazemi (2009).

References

1. Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion contributions). J. Roy. Statist. Soc. Ser. B 36, 192236.
2. Besag, J. (1977). Some methods of statistical analysis for spatial data. Bull. Inst. Int. Stat. 47, 7792.
3. Cox, D.R.(1975) Partial likelihood, Biometrika 62, 269-276.
4. Cox, D.R., Ried, N.(2004) A note on pseudolikelihood constructed from marginal densities, Biometrika 91, 729-737.
5. Davison, A.C. (2003) Statistical Models Cambridge University Press, Cambridge.
6. Lindsay, B.G.(1988) Composite likelihood methods, Contemp. Math. 80, 221-240.
7. Nematollahi A. R. and Davison, A. C. (2009). Penalized compositelikelihood,unpublished.
8. Nematollahi A. R. and Kazemi, M. R. (2009). Pairwise likelihood inferences for the multivariate AR(1) models , Technical Report, Department of Statistics, Shiraz University.
9. Reinsel, G.C. (1997) Elements of multivariate time series analysis, second edition New York, Springer-Verlag.
10. Varin, C. (2008) On composite marginal likelihoods, Advances in Statistical Analysis, 92, 1-28.
11. Varin, C., Vidoni, P. (2006) Pairwise likelihood inference for ordinal categorical time series, Comput. Stat. data Anal. 51, 2365-2373.

Inference on Geometric Extreme Exponential Distribution under Progressive Type-II Censoring

R. Pakyari

Department of Mathematics, Arak University

In this talk, we provide explicit estimator through an approximation of the likelihood equation for the scaled geometric extreme exponential (GE-exponential) parameter based on progressively Type-II censored samples. The bias and mean squared error of the approximate estimator and the MLE were calculated for a wide range of sample sizes and different progressive censoring schemes through a Monte Carlo simulation study. It is observed that the approximate estimator is comparable to the MLE in terms of both absolute bias and mean squared error.

Keywords: Geometric extreme exponential distribution; Progressive Type-II censoring; Monte Carlo simulation; Fisher information.

1. Introduction

The progressive censoring is one of the important sampling techniques that was first introduced by Herd (1956) and its importance in life testing reliability experiments was discussed by Cohen (1963).

The progressive Type-II censoring is as follows: Consider a life test experiment in which n units are placed in the test. At the time of the first failure, R_1 units are randomly removed from the remaining $n - R_1$ units. At the time of the second failure, R_2 units are randomly removed from the remaining $n - R_1 - R_2$ units, and so on. Finally, after the mth failure, R_m remaining units are removed. Thus, we observe m failures and $R_1 + R_2 + \ldots + R_m$ items are progressively censored from the n units, so that $n = m + (R_1 + R_2 + \ldots + R_m)$. The $R = (R_1, \ldots, R_m)$ is called the censoring scheme and is fixed prior to the study. If $R = (0, \ldots, 0)$, no withdrawals are made which correspond to the complete sample and the ordinary order statistics will be obtained. If $R = (0, \ldots, 0, n-m)$, we obtain the conventional Type-II right censoring. We will denote the progressive Type-II censored data by $X_{1:m:n} < X_{2:m:n} < \ldots < X_{m:m:n}$. For an excellent discussion on progressive Type-II censoring technique see the monograph of Balakrishnan and Aggarwala (2000) and Balakrishnan (2007).

There are numerous articles in literature dealing with inferential procedures

based on the progressively Type-II censoring data for a wide variety of lifetime distributions. See for example Mann (1969; 1971), Balakrishnan et al. (2004), Balakrishnan and Kannan (2000), Balakrishnan et al. (2003), Balakrishnan and Asgharzadeh (2005) and Lin et al. (2006) among others who study the inferential procedures for the Weibull, extreme value, logistic, normal, half-logistic and log-gamma models, respectively.

GE-exponential distribution was introduced by Marshal and Olkin (1997) and further studied extensively by Adamidis et al. (2005). The discrimination between the GE-exponential model and the Weibull, gamma and lognormal families as alternatives was studied by Marshal et al. (2001). Recently, Pakyari (2009) studied the discrimination procedures between the GE-exponential and the generalized exponential and Weibull models. He provided the probability of correct selection based on the likelihood ratio and the minimum Kolmogorov distance criteria.

2. Maximum likelihood estimator of the scale parameter

Let the failure times follow a geometric extreme exponential distribution with known shape parameter γ and unknown scale parameter λ and with the probability density function

$$f(x;\lambda) = \frac{\gamma\lambda\, e^{-\lambda x}}{(1-\bar{\gamma}\, e^{-\lambda x})^2}\,, \qquad (x>0) \tag{1}$$

where $\bar{\gamma} = 1-\gamma$.

Let $X_{1:m:n},\ldots,X_{m:m:n}$ be a progressively Type-II censored sample from $f(x;\lambda)$. The likelihood function based on the censored sample is given by

$$\begin{aligned} L(\lambda) = C\, \textstyle\prod_{i=1}^{m} \frac{\gamma\,\lambda\, e^{-\lambda x_i}}{(1-\bar{\gamma} e^{-\lambda x_i})^2}\{1-\frac{1-e^{-\lambda x_i}}{1-\bar{\gamma} e^{-\lambda x_i}}\}^{R_i} \\ C(\gamma\,\lambda)^m \textstyle\prod_{i=1}^{m} \frac{\gamma^{R_i}\, e^{-\lambda(R_i+1)x_i}}{(1-\bar{\gamma}\, e^{-\lambda x_i})^{R_i+2}}\,, \end{aligned} \tag{2}$$

where

$$C = n(n-1-R_1)(n-2-R_1-R_2)\ldots(n-m+1-R_1-\ldots-R_{m-1})\,.$$

The log-likelihood function may be written as

$$L^*(\lambda) = \log L(\lambda) = K + m\,\log\lambda - \lambda\sum_{i=1}^{m}(R_i+1)x_i - \sum_{i=1}^{m}(R_i+2)\log(1-\bar{\gamma}\, e^{-\lambda x_i})\,, \tag{3}$$

where K is a constat. Let define $z_i = \lambda x_i$, now the log-likelihood function may then be rewritten as

$$L^*(\lambda) = K + m\,\log\lambda - \sum_{i=1}^{m}(R_i+1)z_i - \sum_{i=1}^{m}(R_i+2)\log(1-\bar{\gamma}\, e^{-z_i})\,. \tag{4}$$

The likelihood equation may be simplified as

$$m - \sum_{i=1}^{m}(R_i+1)z_i - \sum_{i=1}^{m}(R_i+2)(\frac{\bar{\gamma}z_i e^{-z_i}}{1-\bar{\gamma}e^{-z_i}}) = 0\,. \tag{5}$$

The MLE of the scale parameter λ is the solution of the nonlinear equation (5) which can not be expressed explicitly. In the following section, we present an approximation for the likelihood equation which yields an explicit expression for the MLE of λ.

3. Approximate MLE

Let us approximate the likelihood equation (5) by approximating the function $h(z) = \frac{e^{-z}}{1-\bar{\gamma}e^{-z}}$ in a Taylor series expansion around $E(Z_{i:m:n}) = \nu_{i:m:n}$. Let $U_{i:m:n}$ is the ith order statistic of a progressively Type-II censored sample from the $U(0,1)$ distribution. We then have

$$Z_{i:m:n} = F^{-1}(U_{i:m:n}) = \log(\frac{\bar{\gamma}\,U_{i:m:n}-1}{U_{i:m:n}-1})\,,$$

where $\alpha_{i:m:n} = E(U_{i:m:n})$ and

$$\alpha_{i:m:n} = 1 - \prod_{j=m-i+1}^{m}\frac{j+R_{m-j+1}+\ldots+R_m}{j+1+R_{m-j+1}+\ldots+R_m}\,, \qquad ;i=1,\ldots,m.$$

Expanding $h(z_i)$ around $\nu_{i:m:n}$ by Taylor series expansion and keeping the first two terms, we deduce

$$\begin{aligned} h(z_i) &\approx h(\nu_{i:m:n}) + h'(z_i)|_{z=\nu_{i:m:n}}(z_i-\nu_{i:m:n}) \\ &= \alpha_i + \beta_i z_i\,, \end{aligned} \tag{6}$$

Therefore the likelihood equation may be approximated by

$$\frac{dL^*}{d\lambda} \approx m - \sum_{i=1}^{m}(R_i+1)z_i - \bar{\gamma}\sum_{i=1}^{m}(R_i+2)z_i(\alpha_i+\beta_i z_i) = 0\,, \tag{7}$$

Equation (7) may be rewritten as

$$\left\{\bar{\gamma}\sum_{i=1}^{m}(R_i+2)\beta_i x_i^2\right\}\lambda^2 + \left\{\bar{\gamma}\sum_{i=1}^{m}(R_i+2)\alpha_i x_i + \sum_{i=1}^{m}(R_i+1)x_i\right\}\lambda - m = 0 \tag{8}$$

or

$$A\lambda^2 + B\lambda - m = 0\,, \tag{9}$$

Equation (9) is a quadratic equation in terms of λ with the following two solutions:

$$\frac{-B \pm \sqrt{B^2+4Am}}{2A}$$

However, one of the solutions is inadmissible since $A > 0$. Hence the approximate MLE of the scale parameter λ is given explicitly by

$$\tilde{\lambda} = \frac{-B + \sqrt{B^2 + 4Am}}{2A}\,. \tag{10}$$

4. Fisher Information

In this section, we derive the asymptotic variance of the maximum likelihood estimator of the scale parameter λ for the full and approximate likelihood equations. We shall denote these quantities by $\hat{I}^{-1}$ and $\tilde{I}^{-1}$, respectively.

The observed Fisher information may be obtain from the full likelihood equation as

$$\frac{d^2L^*}{d\lambda^2} = -\frac{m}{\lambda^2} + \frac{1}{\lambda^2}\sum_{i=1}^{m}(R_i+2)\frac{\bar{\gamma}z_i^2e^{-z_i}}{(1-\bar{\gamma}e^{-z_i})^2}\,. \tag{11}$$

Also from the approximate likelihood equation (7) we deduce the observed Fisher information as

$$\frac{d^2L^*}{d\lambda^2} \approx -\frac{1}{\lambda}\sum_{i=1}^{m}(R_i+1)z_i - \frac{\bar{\gamma}}{\lambda}\sum_{i=1}^{m}(R_i+2)z_i(\alpha_i+2\beta_i z_i)\,. \tag{12}$$

The asymptotic variance of the MLE of the scale parameter λ is then given by inverting the observed Fisher information from the full or approximate likelihood equations evaluated at $\lambda = \hat{\lambda}$ or $\lambda = \tilde{\lambda}$, respectively.

We shall use these asymptotic variances on Section 6 to derive confidence interval for the scale parameter λ.

5. Simulation study

In this section, we assess the performance of the proposed approximate estimator and the MLE by a large amount of Monte Carlo simulation experiments. All the simulations were carried out in R using the pseudo-random generator in that software package. We generated the progressively Type-II censored samples from the GE-exponential distribution using the algorithm presented by Balakrishnan and Sadhu (1995).

Five values of the sample sizes say, $n = 15, 20, 30, 40, 50$, different choices of the effective sample size m, and different progressive schemes were considered. For brevity in notation we follow the Balakrishnan (2004) notation to denote the censoring scheme. For example $(2 * 7, 4 * 0)$ denotes the censoring scheme $(7, 7, 0, 0, 0, 0)$.

We computed the approximate MLE from equation (10). The MLE was then calculated from the nonlinear equation (5) by an iterative procedure such as Newton-Raphson method with the approximate estimator as the starting value. We repeated this procedure 10000 times.

Tables 1 and 2 give the average of absolute bias, mean squared error, variance, and variance calculated from observed Fisher information for the approximate estimator and the MLE, respectively. We observe that the approximate MLE is comparable to the MLE in terms of both absolute bias and mean squared error for almost all the sample sizes and the censoring schemes. Also, as one would expect, both the bias and the mean squared error decrease as the effective sample proportion i.e. m/n increases. Moreover, the observed Fisher information tends to be close to the variance of the estimator for both the approximate estimator and the MLE when the effective sample proportion increases.

The tables also show that for all sample sizes n and effective sample sizes m, the censoring scheme $R = (n-m, 0, \ldots, 0)$ tends to the smallest bias and mean squared error, whilst, the censoring scheme $R = (0, 0, \ldots, n-m)$ tends to the largest results.

References

1. Adamidis, K., Dimitrakopoulou, T. and Loukas, S. (2005). On an extension of the exponential-geometric distributions. *Statistics and Probability Letters* **73**, 259-269.
2. Balakrishnan, N. (2007). Progressive censoring methodology: an appraisal. *Test* **16**, 211-259.
3. Balakrishnan, N., Aggrawala, R. (2000). *Progressive Censoring: Theory, Methods, and Applications.* Birkhauser, Boston.
4. Balakrishnan, N., Asgharzadeh, A. (2005). Inference for the scaled half-logistic distribution based on the progressively Type-II censored samples. *Coomunications in Satistics, Theory and Methods* **34**, 73-87.
5. Balakrishnan, N., Kannan, N. (2000). Point and interval estimation for the parameters of the logistic distribution based on progressively Type-II censored samples. In N. Balakrishnan, & C. R. Rao (Eds), *Handbook of Satistics: Vol.* **20**, 431-456.
6. Balakrishnan, N., Kannan, N., Lin, C.T., Ng, H.K.T. (2003). Point and interval estimation for the normal distribution based on progressively Type-II censored samples. *Journal of Satistical Computation and Simulation* **74**, 25-45.
7. Balakrishnan, N., Kannan, N., Lin, C.T., Wu, S.J.S. (2004). Inference for the extreme value distribution under progressive Type-II censoring. *Journal of Satistical Computation and Simulation* **74**, 25-45.
8. Balakrishnan, N., Sandhu, R.A (1995). A simple simulational algorithm for generating progressive Type-II censored samples. *The American Satistican* **49(2)**, 229-230.
9. Cohen, A.C. (1963). Progressively censored samples in life testing. *Technometrics* **5**, 327-329.
10. Erdeli, A., Magnus, W., Oberhettinger, F. and Tricomi, F.G. (1953). *Higher Transcendental Functions.* McGraw-Hill, New York.

11. Herd, R.G. (1956). Estimation of the parameters of a population from a multi-censored sample. PhD Thesis, Iowa State College, Ames, Iowa.
12. Lieblein, J., and Zelen, M. (1956). Statistical investigation of the fatigue life of deep-groove ball bearings. *J. Res. Natl. Bur. Stand.* **57**, 273-316.
13. Lin, C.T., Wu, S.J.S., Balakrishnan, N. (2006). Inference for log-gamma distribution based on progressively Type-II censored data. *Coomunications in Satistics, Theory and Methods.* **35**, 1271-1292.
14. Mann, N.R. (1969). Exact three-order statistic confidence bounds on reliable life for a Weibull model with progressive censoring. *Journal of the American Statistical Association.* **64**, 306-315.
15. Mann, N.R. (1971). Best linear invariant estimation for Weibull parameters under progressive censoring. *Technometrics.* **13**, 521-533.
16. Marshall, A.W. and Olkin, I. (1997). New families of life distributions with extreme stability. *Biometrika* **84**, 641-652.
17. Marshall, A.W., Meza, J.C. and Olkin, I. (2001). Can data recognize its parent distribution? *Journal of Computational and Graphical Statistics* **10**, 555-580.
18. Pakyari, R. (2009). Discriminating between generalized exponential, geometric extreme exponential and Weibull distributions. *Journal of Statistical Computation and Simulation.* DOI: 10.1080/00949650903173306.

n	m	Scheme	AB($\tilde{\lambda}$)	MSE($\tilde{\lambda}$)	var($\tilde{\lambda}$)	$\tilde{I}^{-1}$
15	5	(4*0, 10)	0.1936	0.3022	0.2647	0.1210
15	5	(10, 4*0)	0.1103	0.1925	0.1804	0.0973
15	5	(5*2)	0.1726	0.2621	0.2324	0.1111
15	5	(2*0, 2, 2*4)	0.1881	0.2910	0.2556	0.1180
20	8	(7*0, 12)	0.1082	0.1203	0.1086	0.0687
20	8	(12, 7*0)	0.0592	0.0890	0.0855	0.0592
20	8	(3*4, 5*0)	0.0629	0.0938	0.0899	0.0608
20	8	(5, 4, 3, 5*0)	0.0626	0.0932	0.0893	0.0606
30	10	(9*0, 20)	0.0895	0.0914	0.0834	0.0576
30	10	(20, 9*0)	0.0449	0.0659	0.0638	0.0474
30	10	(9, 7, 4, 7*0)	0.0471	0.0687	0.0664	0.0486
30	10	(10*2)	0.0768	0.0805	0.0746	0.0527
50	15	(14*0, 35)	0.0596	0.0515	0.0479	0.0391
50	15	(35, 14*0)	0.0281	0.0384	0.0372	0.0317
50	15	(10, 9, 8, 7, 1, 10*0)	0.0296	0.0400	0.0392	0.0392
50	15	(11*3, 2, 3*0)	0.0349	0.0436	0.0424	0.0345

n	m	Scheme	AB($\hat{\lambda}$)	MSE($\hat{\lambda}$)	var($\hat{\lambda}$)	$\hat{I}^{-1}$
15	5	(4*0, 10)	0.1930	0.3017	0.2664	0.1807
15	5	(10, 4*0)	0.1348	0.2061	0.1879	0.1464
15	5	(0, 2*5, 2*0)	0.1546	0.2405	0.2167	0.1627
15	5	(5*2)	0.1739	0.2630	0.2328	0.1604
15	5	(2*0, 2, 2*4)	0.1876	0.2906	0.2554	0.1744
20	8	(7*0, 12)	0.1079	0.1201	0.1085	0.0852
20	8	(12, 7*0)	0.0783	0.0949	0.0887	0.0769
20	8	(3*4, 5*0)	0.0829	0.1002	0.0933	0.0802
20	8	(5, 4, 3, 5*0)	0.0824	0.0995	0.0927	0.0797
30	10	(9*0, 20)	0.0892	0.0913	0.0834	0.0690
30	10	(20, 9*0)	0.0615	0.0696	0.0658	0.0588
30	10	(9, 7, 4, 7*0)	0.0644	0.0727	0.0685	0.0609
30	10	(10*2)	0.0793	0.0812	0.0749	0.0628
50	15	(14*0, 35)	0.0594	0.0514	0.0479	0.0441
50	15	(35, 14*0)	0.0399	0.0396	0.0380	0.0366
50	15	(10, 9, 8, 7, 1, 10*0)	0.0423	0.0418	0.0399	0.0384
50	15	(11*3, 2, 3*0)	0.0498	0.0457	0.0433	0.0408

Improving On The Minimum Risk Equivariant Estimator In The Spherically Symmetric Distributions

A.T. Payandeh

Department of Mathematical Sciences, Shahid Beheshti University

We propose improvements under quadratic loss function to the minimum risk equivariant (mre) estimator for the estimating the mean of a p-variate spherically symmetric distribution, when unknown mean lies in a ball of radius m centered at the origin and the covariance matrix $\sigma_0^2 I_p$ is given. Our construction of explicit improvements relies on Kubokawa's Integral Expression of Risk Difference (IERD) method. For large class of scale-mixture of normal distributions, a dominator for the mre will be addressed.

Keywords: Spherical symmetric distribution,; Integral Expression of Risk Difference method; Dominator estimator; Quadratic loss function.

1. Introduction

Consider problem of estimating, under quadratic loss, the location parameter θ of a spherically symmetric distribution, basis on an observation X. For nonrestricted parameter space, Stein in (1955) investigated the question of admissibility of the minimum risk equivariant estimator (mre), $\delta_0(x) = x$, when sampling from a multinormal distribution with respect to quadratic loss. Furthermore, in 1961 James and Stein proved (**i**) inadmissibility of the mre is not exclusive to case of quadratic loss but is true for any loss functions $L(\delta,\ \theta) = s(||\delta - \theta||)$, where s is a continuous, differentiable and concave function and (**ii**) the results do not restricted to normal distribution but are true for other family of distributions, with some certain conditions. One might infer as well from this work that inadmissibility of the mre, when the mean vector has 3 or more dimensions, could be generalized to a large class of distributions. Brown (1966) considered the general location parameter estimation problem and proved that under mild assumptions on the loss function, the mre is admissible in one and two dimensions and inadmissible in three or more dimensions. The results of Stein and Brown suggest a new problem. Namely, finding explicit estimators which are performance better, in sense of risk function, than the mre.

Baranchik (1964) found a class of minimax estimators for multivariate normal mean, which includes the James-Stein class of estimators. After Baranchik's discovery, until 1974, explicit estimators which perform better than the mre were only available for mean vector of multivariate normal distribution. Then, Strawderman (1974) and Berger (1975) found minimax estimator which per-

form better than the mre, when X sample from a certain class of spherically symmetric distributions. Cohen Brandwein & Strawderman (1978) showed that $\delta_{a,g}(x) = (1 - ag(||x||^2/||x||^2)x$, under some conditions on g and a, performs better than the mre. More recently, their results improved by Xu & Izmirlian (2006).

The above findings don't be valid anymore, whenever the parameter space has been constrained, in some sense. In such situation, the mre *not only* inadmissible *but also* with positive probability take its value outside of parameter spaces. Casella & Strawderman (1981) established that the boundary uniform Bayes estimator, given by $\delta_{BU}(x) = m\tan(mx)$ is minimax, whenever the underlying distribution is univariate normal and the location parameter, θ is bounded by an interval $[-m, m]$, and $m \leq 1.0567$. Casella & Strawderman's result generalized to wider truncated parameter spaces using N-point prior by Gatsonis, MacGibbon, and Strawderman in (1987). Berry (1990) and Marchand & Perron (2002) extended Casella & Strawderman discovery to multivariate normal distribution and by DasGupta (1985) to a wide class of multivariate distributions. Hartigan (2004) discovered that, for multinormal distribution, the fully uniform Bayes procedure for a wide class of constrained parameter spaces, such as Ball, Cone and etc, is better than the mre.

That seem finding dominating estimator for maximum likelihood estimator (mle), in general, is very difficult. Several authors considered problem of finding dominator for the mle whenever the parameter space has been constrained. A good review can be found in Marchand & Strawderman (2004) and van Eeden (2006), among others. As far as we know, no any Bayes estimator available whose improves the mle for any class of constrained parameter space. For this reason that make sense to find an improvement estimator for the mre rather than the mle.

Kubokawa in a series of papers established a technique named Integral Expression of Risk Difference (IERD) method to find dominator for a given estimator. The IERD's method has been explained in Marchand & Strawderman (2004). Marchand & Strawderman (2005) employed IERD's method to construct a class of dominating estimator for the mre for a class univariate location family distributions, under a quite general class of loss functions. Kubokawa (2005) benefited from Marchand & Strawderman (2005)'s results and found a class of dominating Bayes estimator for the mre, under the squared-error loss. The discovery of Marchand & Strawderman (2005) inspired us to develop their results to the multivariate spherically symmetric distributions, under the quadratic loss function.

This paper, using the IERD's method, develops Hartigan (2004)'s discovery to a wide class of spherically symmetric distributions. This paper developed as the following. Section 2 derives a general sufficient condition to an estimator with form $\delta_g(x) = g(||x||)x$ dominates the mre. Sections (3) for multinormal distribution constructs a dominator procedure for the liner minimax estimator and consequently for the mre. Finally, Section (4), for a wide class of scale-mixture normal distributions presents an improvements for the mre.

2. Main results

Throughout, we work with loss function of the form $L(\theta, d) = ||d - \theta||^2$ and a spherically symmetric family of distribution with density function

$$f_\theta(x) \propto k(\frac{||x-\theta||^2}{\sigma_0^2}), \tag{1}$$

where unknown vector mean θ is constrained in ball with radius m, say Θ_m, σ_0^2 is given scale parameter, and k is a symmetric function about zero. From here now, we use notions $R = ||X||$, $r = ||x||$, and $\lambda = ||\theta||$, where $||y||$ stands for norm of y.

It is useful to recall that: A family of densities function $\{p_\theta(\cdot) : \theta \in \Theta\}$ is said to have monotone likelihood ratio (mlr) in $T(\cdot)$, such that for all $\theta_1 > \theta_2$, the densities $p_{\theta_i}(\cdot)$, for $i = 1, 2$, are distinct, and the ratio $p_{\theta_1}(x)/p_{\theta_2}(x)$ is a nondecreasing function of $T(x)$.

Now, we present the main theorem of this paper.

Theorem 2.1. *Suppose random variable X sampled from a spherically symmetric family of distributions 1. Then, under the quadratic loss $\delta_g(X) = g(||X||)X$ dominates the mre, $\delta_0(X) = X$, whenever **(i)** g absolutely continuous and nonincreasing; **(ii)** $g(r) \geq sup_{\theta\in\Theta_m} H(r, \theta)$; **(iii)** $g(0) \in [m^2/(m^2 + p),\ 1]$, where $H(r, \theta) = E_\theta(\theta^T X |\ ||X|| \geq r)/E_\theta(X^T X |\ ||X|| \geq r)$.*

Proof. The desire proof arrives by showing that: Under (**i**) and (**ii**) $\delta_g(x)$ dominates $g(0)x$ and under (**iii**) $g(0)x$ dominates $\delta_0(x)$. Now, using the IERD's method, risk difference between two estimators $\delta_g(x)$ and $g(0)x$ can be written as

$$\begin{aligned}
\frac{1}{2}\Delta(\theta) &= \frac{1}{2}\left[R(\theta, g(||X||)X) - R(\theta, g(0)X)\right] \\
&= \frac{1}{2}\left[E_\theta||g(||X||)X - \theta||^2 - ||g(0)X - \theta||^2\right] \\
&= \frac{1}{2}E_\theta\left(\int_0^{||X||} \frac{\partial}{\partial t}||g(t)X - \theta||^2 dt\right) \\
&= \int_0^\infty \int_0^r g'(t)[g(t)x - \theta]^T x f_p(r, \lambda) dt dr \\
&= \int_0^\infty g'(t) \int_t^\infty [g(t)x^T x - \theta^T X] f_p(r, \lambda) dr dt,
\end{aligned}$$

where $f_p(r, \lambda)$ is density of $R = ||X||$. Now, from (**i**) and (**ii**) observe the second integrand is positive. This observation establishes that $\delta_g(x)$ dominates $g(0)x$. To complete the desire proof, one has to show $g(0)x$ dominates $\delta_0(x)$. Difference between risk of $g(0)x$ and $\delta_0(x)$ can be explored as

$$\begin{aligned}
\Delta(\theta) &= R(\theta, g(0)X) - R(\theta, X) \\
&= E_\theta(||g(0)X - \theta||^2 - ||X - \theta||^2) \\
&= E_\theta(g^2(0)X^T X + ||\theta||^2 - 2g(0)\theta^T X - p) \\
&= (g(0) - 1)^2\lambda^2 + (g^2(0) - 1)p \\
&\leq (g(0) - 1)^2 m^2 + (g^2(0) - 1)p \leq 0.
\end{aligned}$$

The last expression follows from (**iii**) completes the desire proof. □

The above theorem provides a sufficient condition to generate dominators for δ_0. The following provides a weaker condition to generate dominators for δ_0.

Corollary 2.1. *Suppose random variable X sampled from a spherically symmetric family of distributions 1. Then, condition(**ii**) in Theorem (2.1) can be replaced by a weaker condition $g(r) \geq sup_{\theta\in\Theta_m} H^*(r,\theta)$, where $H^*(r,\lambda) = \lambda E_\theta(R|R \geq r)/E_\theta(R^2|R \geq r)$.*

Proof. Application of the Cauchy-Schwartz inequality, i.e., $||\theta^T X|| \leq ||\theta||||X||$, leads to the desire proof. □

The next lemma provides the linear minimax estimator, say δ_{LME}, for a class of spherically symmetric distributions.

Lemma 2.1. *Suppose X distributed according to a spherical symmetric distribution 1 with an unknown location parameter θ, where restricted in a ball with radius m. Then, the linear minimax estimator δ_{LME} (**i**) is given by $\delta_{LME}(x) = m^2x/(m^2 + p\sigma_0^2)$; (**ii**) dominates δ_0.*

Proof. Consider a class of linear estimators for θ with form $D_L = \{\delta_a : \ \delta_a(x) = ax; \ \ a \in \mathbb{R}^+\}$. By maximizing risk $\delta_a(x)$, in D_L, with respect to λ then minimizing risk with respect to a, the linear minimax estimator, given above, arrives. To establish (**ii**), set $a_{opt} = m^2/(m^2 + p\sigma_0^2)$. Now observe that, difference between risks δ_{LME} and δ_0 can be stated as

$$\begin{aligned}\Delta(\theta) &= R(\theta, \delta_{LME}(X)) - R(\theta, X)\\ &= (a_{opt}^2 - 1)p + (a_{opt} - 1)^2\lambda^2\\ &\leq (a_{opt}^2 - 1)p + (a_{opt} - 1)^2m^2\\ &= (a_{opt} - 1)\frac{p^2 + mp}{m^2 + p} \leq 0,\end{aligned}$$

where the last inequality follows from the fact $0 \leq a_{opt} \leq 1$. □

Two estimators $\delta_H(x) = H(||x||, m)x$ and $\delta_{H^*}(x) = H^*(||x||, m)x$ can be employed to obtain a dominator for δ_0 and δ_{LME}, respectively. The following explores some properties of δ_H and δ_{H^*}.

Lemma 2.2. *Consider bivariate functions $H(\cdot,\cdot)$ and $H^*(\cdot,\cdot)$, given by Theorem 2.1 and Corollary 2.1, respectively. Then, (**i**) Two estimators $\delta_H(x) = H(r,m)x$ and $\delta_{H^*}(x) = H^*(r,m)x$ are range-preserving, whenever $H(\cdot,\lambda)$ and $H^*(\cdot,\lambda)$ are increasing functions in λ; (**ii**) δ_H and dominates δ_{LME} and δ_0, whenever $H(r,\lambda)$ and $H^*(r,\lambda)$ are increasing in λ and decreasing in r; (**iii**) δ_{H^*} dominates δ_0, whenever $H(r,\lambda)$ and $H^*(r,\lambda)$ are increasing in λ and decreasing in r.*

Proof. Since $H(\cdot,\lambda)$ and $H^*(\cdot,\lambda)$ are increasing functions in λ. Then, $H(\cdot,m) = sup_{\theta\in\Theta_m} H(\cdot,\theta)$ and $H^*(\cdot,m) = sup_{\theta\in\Theta_m} H^*(\cdot,\theta)$. For part (**i**), from the fact that $||H^*(r,m)|| \leq m$ observes

$$H^*(r,m) = m\frac{\int_r^\infty y f_p(y,m)dy}{\int_r^\infty y^2 f_p(y,m)dy} \leq \frac{m}{r}.$$

Range-preserving of δ_H follows from the above observation and the fact that $H(r,m) < H^*(r,m)$. For part (**ii**), from Theorem 2.1 (parts i and ii) observe that $\delta_H(x) = H(r,m)x$ dominates $H(0,m)x = \delta_{LME}(x)$ and from Lemma 2.1 (part ii) it dominates δ_0. An estimator $\delta_{H^*}(x) = H^*(r,m)$ satisfies conditions given by Theorem 2.1. Therefore, δ_{H^*} dominates δ_0. This observation complete the desire proof. □

The proceeding sections consider some spacial class of spherical symmetric distributions and employ results of this section to obtain dominator for δ_0 and δ_{LME}.

3. Multinormal distribution

Suppose random variable X is distribution with a p-variate normal distribution with unknown location parameter θ and given covariance matrix $\sigma_0^2 I_p$. The modified Bessel function $\mathbb{I}_v(x)$ and $\rho_v(x) = \mathbb{I}_{v+1}(x)/\mathbb{I}_v(x)$ play a key roles in the rest of paper. The following due to Watson (1983) and Amos (1974) explores some useful properties for $\mathbb{I}_p$ and ρ_p.

Lemma 3.1. *For $p \geq 1$,*

i) $\mathbb{I}_{p-1}(x) = \frac{2p}{x}\mathbb{I}_p(x) + \mathbb{I}_{p+1}(x)$;
ii) $\rho_{p/2-1}(x)$ *is increasing, concave, and* $\lim_{x\to\infty} \rho_{p/2-1}(x) = 1$;
iii) $\rho_{p/2-1}(x)/x$ *is decreasing in x with* $\lim_{x\to 0} \rho_{p/2-1}(x)/x = 1/p$;
iv) $\rho_{p/2}(x) = \rho^{-1}_{p/2-1} - p/x$.

The next lemma, due to Marchand & Perron (2001), collects some well known properties of multinormal distribution.

Lemma 3.2. *Suppose X is distributed according to a $N_p(\theta, \sigma_0^2 I_p)$. Then,*

i) $f_p(r,\lambda) = r(\frac{r}{\lambda})^{p/2-1}\mathbb{I}_{p/2-1}(r\lambda)exp\{-\frac{r^2+\lambda^2}{2}\}$,
ii) $r^2 f_p(r,\lambda) = \lambda^2 f_{p+4}(r,\lambda) + p f_{p+2}(r,\lambda)$,
iii) $\rho(\lambda, y) = \lambda\rho_{p/2-1}(r\lambda)$,
iv) $r f_p(r,\lambda) = \lambda f_{p+2}(r,\lambda)$,
v) $2\frac{\partial}{\partial\lambda}F_p(r,\lambda) = F_{p+2}(r,\lambda) - F_p(r,\lambda)$,

where $f_p(r,\lambda)$ is density function of $R = ||X||$, $\rho(\lambda, y) = E_\theta(\frac{\theta^T X}{Y} |\ Y = y)$, and F is the cumulative distribution function, CDF, of random variable R.

The following theorem provides a dominator for the linear minimax estimator δ_{LME}, and consequently for δ_0.

Theorem 3.1. *Suppose X has been distributed according to a $N_p(\theta, \sigma_0^2 I_p)$, where an unknown mean vector θ is restricted in a ball with radius m. Then, $\delta_H(x) = H(r,m)x$ dominates the linear minimax estimator $\delta_{LME}(x) = m^2 x/(m^2 + p\sigma_0^2)$, where $H(r,\lambda) = E_\theta(\theta^T X |\ ||X|| \geq r)/E_\theta(X^T X |\ ||X|| \geq r)$.*

Proof. It suffices to show that $H(r,\lambda)$ is increasing in λ and nonincreasing in r. The rest of proof follows from Lemma 2.2. $H(r,\lambda)$ may be understood as

$$H(r,\lambda) = \frac{E_\theta(||X|| E_\theta^{||X||}(\frac{\theta^T X}{||X||}|\ ||X|| \geq r))}{E_\theta(||X||^2|\ ||X|| \geq r)} = \frac{\int_r^\infty y\lambda\rho_{p/2-1}(\lambda y) f_p(y,\lambda)dy}{\int_r^\infty y^2 f_p(y,\lambda)dy} = \frac{\int_r^\infty \lambda^2 f_{p+2}(y,\lambda)dy}{\int_r^\infty y^2 f_p(y,\lambda)dy}.$$

Using Lemma (3.2), one can observe

$$\frac{1}{H(r,\lambda)} = \frac{p}{\lambda^2} + \frac{\int_r^\infty f_{p+4}(y,\lambda)dy}{\int_r^\infty f_{p+2}(y,\lambda)dy} = \frac{p}{\lambda^2} + \frac{1-F_{p+4}(r,\lambda)}{1-F_{p+2}(r,\lambda)}$$

Derivative of the second term, with respect to λ, is proportional to

$$\begin{aligned} &- [F_{p+6}(y,\lambda) - F_{p+4}(y,\lambda)]\,[1-F_{p+2}(y,\lambda)] \\ &\quad + [F_{p+4}(y,\lambda) - F_{p+2}(y,\lambda)]\,[1-F_{p+4}(y,\lambda)] \\ &\quad = \overline{F}_{p+6}(y,\lambda)\overline{F}_{p+2}(y,\lambda) - \overline{F}^2_{p+4}(y,\lambda) \leq 0, \end{aligned}$$

which the last inequality follows from properties of noncentral chi-square. This observation established that $H(\cdot,\lambda)$ increasing in λ. To show $H(r,\cdot)$ decreasing in r, the following presentation of $H(r,\lambda)$ is useful to consider.

$$H(r,\lambda) = E_{r,\lambda}(\frac{E_Y(\frac{\theta^T X}{||X||}|\ ||X|| = Y)}{Y}) = E_{r,\lambda}(\frac{\rho_{p/2-1}(\lambda Y)}{Y}),$$

where Y has density proportional to $f_p(y,\lambda)1_{[r,\infty)}(y)$. Obviously, density function of Y has the MLR in y, whenever r viewed as a parameter. On the hand, the inner expression is decreasing function in r. Now application of the kline-Rubin theorem, Lehmann & Romano(2005), establishes that $H(r,\lambda)$ is a decreasing function in r. □

4. Scale-mixture Normals

Suppose $X|Z = z$ has normal distribution and mixing random variable Z has density function g, which defined with positive measure μ on $(0,\infty)$. One can readily observe that the random vector X has density function

$$f_\theta(x) = \int_0^\infty (2\pi z)^{-p/2} exp\{-\frac{||x-\theta||^2}{2z}\} g(z)\mu(dz). \tag{2}$$

Due to Marchand (1993) the following shows some useful properties of scale-mixture normals.

Lemma 4.1. *Suppose $X|Z = z$ has a $N_p(\theta, zI_p)$, where the unknown vector mean θ is redistricted in ball with radius m. Moreover, random variable Z has been*

distributed according to density function g, with positive measure μ on $(0,\infty)$. Then, density function of random variable $R = ||X||$ is

$$f_p(r,\lambda) = \int_0^\infty \frac{r}{z}(\frac{r}{\lambda})^{p/2-1}\mathbb{I}_{p/2-1}(\frac{\lambda r}{z})exp\{-\frac{r^2+\lambda^2}{2z}\}g(z)\mu(dz). \quad (3)$$

The next theorem constructs a dominating estimator for δ_0.

Theorem 4.1. *Suppose $X|Z = z$ has been distributed according to a $N_p(\theta, zI_p)$), where θ is restricted in a ball with radius m. Moreover, suppose random variable Z defined with positive measure μ on $(0,\infty)$. Then, either one the following two conditions are sufficient to $\delta_{H^*}(x) = H^*(r,m)x$ dominate δ_0.*

i) *Z is distributed uniformly, discretely or continuously, on any subinterval of $(0,\infty)$;*
ii) *random variable Z has a log-concave density function.*

Proof. It suffices to show that $H^*(r,\lambda)$ nonincreasing in r and, under either (i) or (ii), increasing in λ. The rest of proof follows from Lemma 2.2. $H^*(r,\cdot)$ decreasing, because

$$\frac{\partial}{\partial r}H^*(r,\lambda) = \frac{\lambda r f_p(r,\lambda)}{(\int_r^\infty y^2 f_p(r,\lambda)dy)^2}\int_r^\infty y(r-y)f_p(y,\lambda)dy \le 0,$$

where f_p is density of R, given by Equation 3 and the last inequality follows from fact that $y \ge r$.

From definition of H^*, one can observe that

$$\begin{aligned} H^*(r,\lambda) &= \frac{\lambda E_\theta(R|R\ge r)}{E_\theta(R^2|R\ge r)} \\ &= \frac{\lambda\int_r^\infty\int_0^\infty y^{\frac{p}{2}}\frac{g(z)}{z}\mathbb{I}_{\frac{p}{2}-1}(\frac{y\lambda}{z})e^{-\frac{y^2+\lambda^2}{z}}\mu(dz)dy}{\int_r^\infty\int_0^\infty y^{\frac{p}{2}+1}\frac{g(z)}{z}\mathbb{I}_{\frac{p}{2}-1}(\frac{y\lambda}{z})e^{-\frac{y^2+\lambda^2}{z}}\mu(dz)dy} \\ &= \frac{\int_{r/\lambda}^\infty\int_0^\infty x^{\frac{p}{2}}\frac{g(\lambda^2 t)}{t}\mathbb{I}_{\frac{p}{2}-1}(\frac{x}{t})e^{-\frac{1+x^2}{t}}\mu(dt)dx}{\int_{r/\lambda}^\infty\int_0^\infty x^{\frac{p}{2}+1}\frac{g(\lambda^2 t)}{t}\mathbb{I}_{\frac{p}{2}-1}(\frac{x}{t})e^{-\frac{1+x^2}{t}}\mu(dt)dx}, \end{aligned}$$

where the last equation obtained after transforming $y = \lambda x$ and $z = \lambda^2 t$. Derivative of H^*, with respect to λ, is $\frac{\partial}{\partial\lambda}H^*(r,\lambda) = \frac{1}{B^2}\{A_1 - A_2 + A_3 - A_4\}$, where B is

denominator of H^* and

$$A_1 = 2\lambda \int_{r/\lambda}^{\infty}\int_0^{\infty} xM(x,t)\mu(dt)dx \int_{r/\lambda}^{\infty}\int_0^{\infty} x^{\frac{p}{2}} g'(\lambda^2 t)\mathbb{I}_{\frac{p}{2}-1}(\frac{x}{t})e^{-\frac{1+x^2}{t}}\mu(dt)dx;$$

$$A_2 = 2\lambda \int_{r/\lambda}^{\infty}\int_0^{\infty} M(x,t)\mu(dt)dx \int_{r/\lambda}^{\infty}\int_0^{\infty} x^{\frac{p}{2}+1} g'(\lambda^2 t)\mathbb{I}_{\frac{p}{2}-1}(\frac{x}{t})e^{-\frac{1+x^2}{t}}\mu(dt)dx;$$

$$A_3 = \frac{r}{\lambda^2} \int_{r/\lambda}^{\infty}\int_0^{\infty} xM(x,t)\mu(dt)dx \int_0^{\infty} \frac{g(\lambda^2 t)}{t}(\frac{r}{\lambda})^{\frac{p}{2}}\mathbb{I}_{\frac{p}{2}-1}(\frac{r}{\lambda t})e^{-\frac{\lambda^2+r^2}{\lambda^2+t}}\mu(dt);$$

$$A_4 = \frac{r}{\lambda^2} \int_{r/\lambda}^{\infty}\int_0^{\infty} M(x,t)\mu(dt)dx \int_0^{\infty} \frac{g(\lambda^2 t)}{t}(\frac{r}{\lambda})^{\frac{p}{2}+1}\mathbb{I}_{\frac{p}{2}-1}(\frac{r}{\lambda t})e^{-\frac{\lambda^2+r^2}{\lambda^2+t}}\mu(dt);$$

and $M(x,t) = \frac{g(\lambda^2 t)}{t} x^{\frac{p}{2}}\mathbb{I}_{\frac{p}{2}-1}(\frac{x}{t})e^{-\frac{1+x^2}{t}}$. Obviously, $A_3 - A_4 \geq 0$, because $x \geq \frac{r}{\lambda}$. Under condition (**i**) terms A_1 and A_2 will be vanished. Under condition (**ii**), an expression $A_1 - A_2$ can be rewritten as

$$\begin{aligned} A_1 - A_2 = 2\lambda &\int_{r/\lambda}^{\infty}\int_0^{\infty} M(x,t)dtdx \int_{r/\lambda}^{\infty}\int_0^{\infty} h'(\lambda^2 t)xtM(x,t)dtdx \\ &- 2\lambda \int_{r/\lambda}^{\infty}\int_0^{\infty} xM(x,t)dtdx \int_{r/\lambda}^{\infty} h'(\lambda^2 t)tM(x,t)dtdx. \end{aligned}$$

Now, the FKG's inequality (see Fortuin, Kasteleyn, & Ginibre, 1971) states that an expression $A_1 - A_2$ is nonnegative, whenever $M(x_1,t_2)M(x_2,t_1) - M(x_1,t_1)M(x_2,t_2) \leq 0$, for $0 \leq x_1 \leq x_2$ and $0 \leq t_1 \leq t_2$. Substituting M leads to

$$\begin{aligned} M(x_1,t_2)M(x_2,t_1) &- M(x_1,t_1)M(x_2,t_2) \\ &= \mathbb{I}_{\frac{p}{2}-1}(\frac{x_1}{t_2})\mathbb{I}_{\frac{p}{2}-1}(\frac{x_2}{t_1}) - \mathbb{I}_{\frac{p}{2}-1}(\frac{x_1}{t_1})\mathbb{I}_{\frac{p}{2}-1}(\frac{x_2}{t_2}) \\ &\qquad \times exp\{-\frac{x_1^2 - x_2^2}{t_1} - \frac{x_2^2 - x_1^2}{t_2}\} \leq 0, \end{aligned}$$

The above inequality arrives from a double implementation of the Ross's inequality (see Joshi & Bissu, 1991). To observe this, use the Ross's inequality $\frac{\mathbb{I}_{p/2-1}(x_1/t_2)}{\mathbb{I}_{p/2-1}(x_1/t_1)} \leq (t_1/t_2)^{p/2-1}$ and $\frac{\mathbb{I}_{p/2-1}(x_2/t_2)}{\mathbb{I}_{p/2-1}(x_2/t_1)} \geq (t_1/t_2)^{p/2-1} exp\{x_2/t_2 - x_2/t_1\}$, hence $\frac{\mathbb{I}_{p/2-1}(x_1/t_2)}{\mathbb{I}_{p/2-1}(x_1/t_1)} \leq \frac{\mathbb{I}_{p/2-1}(x_2/t_2)}{\mathbb{I}_{p/2-1}(x_2/t_1)} exp\{-(x_1^2 - x_2^2)(1/t_1 - 1/t_2)\}$. Because $0 \leq -(x_1^2 - x_2^2)(1/t_1 - 1/t_2) + x_2(1/t_1 - 1/t_2)$. □

Acknowledgement

This paper has been started jointly with professor Éric Marchand. Author world like to deeply appreciate him for his contributions on early results of paper (especially on Theorem 2.1 and Lemma 2.1) and his constrictive comments on presentation of the final draft. Useful comments of professors Kucerovsky and Arellano-Valle, respectively, on the Ross' inequality and properties of H function (in Theorem 2.1) are deeply appreciated.

References

1. Amos, D. E. (1974). Computation of modified Bessel functions and their ratios. *Mathematics of Computation*, **28**, 239–251.
2. Baranchik, A. J. A (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.*, **41**, 642–645. 62.30
3. Berger, J. (1975). Minimax estimation of location vectors for a wide class of densities. *The Annals of Statistics*, **3**, 1318–1328.
4. Berry, C. (1990). Minimax estimation of a bounded normal mean vector. *Journal of Multivariate Analysis*, **35**, 130–139.
5. Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *The Annals of Mathematical Statistics*, **37**, 1087–1136.
6. Casella, G. & Strawderman, W.E. (1981). Estimating a bounded normal mean. *The Annals of Statistics*, **9**, 870–878.
7. Cohen Brandwein, A. R. & Strawderman, W. E. (1978). Minimax estimation of location parameters for spherically symmetric unimodal distributions under quadratic loss. *Ann. Statist.*, **6**, 377–416.
8. DasGupta, A. (1985). Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. *Sankhya A*, **47**, 326–332.
9. Fortuin, C. M., Kasteleyn, P. W., & Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, **22**, 89–103.
10. Gatsonis, C., MacGibbon, B. & Strawderman, W. E. (1987). On the estimation of a restricted normal mean. *Statistics & Probability Letters*, **6**, 21–30.
11. Hartigan, J. (2004). Uniform priors on convex improve risk, *Statistics & probability letters*, **67**, 285–288.
12. James, W. & Stein, C. (1961). Estimation with quadratic loss. *Procedure of 4th Berkeley Symposium Mathematics, Statistics, and Probability.*, **I**, 361–379.
13. Joshi, C. M. & Bissu, S. K. (1991). Some inequalities of Bessel and modified Bessel functions. *Australian Mathematical Society. Journal. Series A. Pure Mathematics and Statistics*, **50**, 333–342.
14. Kubokawa, T. (2005). Estimation of bounded location and scale parameter. *Journal of the Japan Statistical Society*, **35**, 221–249.
15. Marchand, É . (1993). Estimation of a multivariate mean with constrain on the norm. *The Canadian Journal of Statistics*, **21**, 359–366.
16. Lehmann, E. L. & Romano, J. P. (2005). *Testing statistical hypotheses.* Third edition. Springer Texts in Statistics. Springer, New York.
17. Marchand, É & Perron, F. (2001). Improving on the MLE of a bounded normal mean. *The annals of Statistics*, **29**, 1078–1093.
18. Marchand, É & Perron, F. (2002). On the minimax estimator of a bounded normal mean. *Statistics & Probability Letters*, **58**, 327–333.
19. Marchand, É & Strawderman, W. E. (2004). Estimation in restricted parameter spaces: a review. *A festschrift for Herman Rubin, IMS Lecture Notes Monogr. Ser.*, **45**, 21–44.
20. Marchand, É & Strawderman, W. E. (2005). Improving on the minimum risk equivariant estimator for a location parameter which is constrained to an interval or a half-interval. *Annals of the Institute of Statistical Mathematics*, **57**, 129–143.
21. Stein, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math.*

Statist., **26**, 518–522.

22. Strawderman, W. E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *Journal of Multivariate Analysis*, **4**, 255–264.
23. van Eeden, C. (2006). *Restricted parameter space estimation problems. Admissibility and minimaxity properties.* Lecture Notes in Statistics, 188. Springer, New York.
24. Watson, G. S. (1983). *Statistics on Spheres.* John Wiley & Sons, Inc., New York,
25. Xu, J. L. & Izmirlian, G. (2007). Estimation of location parameters for spherically symmetric distributions. *Journal of Multivariate Analysis*, **97**, 514–525.

Estimation of $P[Y < X]$ for Generalized Exponential Distribution in the Presence of Outliers when Scale Parameter is Known

H. Pazira and P. Nasiri

Department of Statistics, Payame Noor University of Tehran

This paper deals with the estimation of $P(Y < X)$ where Y has generalized exponential distribution with parameters α and λ and X has generalized exponential distribution with presence of k outliers with parameters β_1, β_2 and λ such that X and Y are independent. When scale parameter (λ) is known the maximum likelihood estimator of $R = P(Y < X)$ is derived. Analysis of the simulated and real life data sets have also been presented for illustrative purposes. some of the previous results in the literatures such as Kundu and Gupta (2005) and Nasiri and Jabbari Nooghabi (2009) can be achieved as special case of our results.

Keywords: Stress-Strength model, Maximum likelihood estimator, Outliers.

1. Introduction

Recently the two-parameter generalized exponential (GE) distribution has been proposed by the authors. It has been studied extensively by Gupta and Kundu (1999, 2001a, 2001b, 2002, 2003a, 2003b, 2004, 2006), Raqab (2002), Raqab and Ahsanullah (2001) and Zheng (2002). Note that the generalized exponential distribution is a submodel of the exponentiated weibull distribution introduced by Mudholkar and Srivastava (1993) and later studied by Mudholkar, Srivastara and Freimer (1995) and Mudholkar and Huston (1996).

The two-parameter GE distribution has the following density function

$$f(x, \alpha, \lambda) = \alpha\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{\alpha-1}, \qquad x > 0, \tag{1}$$

and the distribution function

$$F(x, \alpha, \lambda) = (1 - e^{-\lambda x})^{\alpha}, \qquad x > 0. \tag{2}$$

Here $\alpha > 0$ and $\lambda > 0$ are the shape and scale parameters respectively. For different values of the shape parameter, the density function can take different shape. For detailed description of the distribution, one is referred to the original paper of Gupta and Kundu (1999). From now on GE distribution with the shape parameter α and scale parameter λ will be denoted by $GE(\alpha, \lambda)$.

Let the random variables $X_1, X_2, ..., X_{n-k}$ are independent, each having the probability density function $f_1(x)$,

$$f_1(x, \beta_2) = \beta_2 e^{-x}(1 - e^{-x})^{\beta_2 - 1}, \qquad x > 0, \tag{3}$$

and k random variables (as outliers) are also independent, have the probability density function $f_2(x)$,

$$f_2(x, \beta_1) = \beta_1 e^{-x}(1 - e^{-x})^{\beta_1 - 1}, \qquad x > 0. \tag{4}$$

The joint density of $X_1, X_2, ..., X_n$ is given as

$$\begin{aligned}
&f(x_1, x_2, ..., x_n) \\
&= C \prod_{i=1}^{n} f_1(x_i) \cdot \sum_{\underline{A}} \prod_{r=1}^{k} \frac{f_2(x_{A_r})}{f_1(x_{A_r})} \\
&= C\ {\beta_2}^n e^{-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} (1 - e^{-x_i})^{\beta_2 - 1} \sum_{\underline{A}} \frac{{\beta_1}^k e^{-\sum_{r=1}^{k} x_{A_r}} \prod_{r=1}^{k} (1 - e^{-x_{A_r}})^{\beta_1 - 1}}{{\beta_2}^k e^{-\sum_{r=1}^{k} x_{A_r}} \prod_{r=1}^{k} (1 - e^{-x_{A_r}})^{\beta_2 - 1}} \\
&= C\ {\beta_1}^k {\beta_2}^{n-k} e^{-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} (1 - e^{-x_i})^{\beta_2 - 1} \sum_{\underline{A}} \prod_{r=1}^{k} (1 - e^{-x_{A_r}})^{\beta_1 - \beta_2}, x > 0,
\end{aligned} \tag{5}$$

where $C = \frac{k!(n-k)}{n!}$ and $\sum_{\underline{A}} = \sum_{A_1=1}^{n-k+1} \sum_{A_2=A_1+1}^{n-k+2} \cdots \sum_{A_k=A_{k-1}+1}^{n}$. For more details see Dixit (1989), Dixit et al. (1996) and Dixit and Nasiri (2001).

From equation (5), the marginal distribution of X is,

$$h(x, \beta_1, \beta_2) = \frac{k}{n} \beta_1 e^{-x}(1 - e^{-x})^{\beta_1 - 1} + \frac{n-k}{n} \beta_2 e^{-x}(1 - e^{-x})^{\beta_2 - 1}, \qquad x > 0. \tag{6}$$

The main aim of this paper is to focus on the inference of $R = P(Y < X)$, where $Y \sim GE(\alpha, \lambda)$ and X has mixture GE or marginal distribution of $X_1, X_2, ..., X_n$ with presence of k outliers. For Simplify we consider $\lambda = 1$. The estimation of R is very common in the statistical literature. For example, if X is the strength of a component which is subject to a stress Y, when R is a measure of system performance and arises in the context of mechanical reliability of a system. We obtain the maximum likelihood estimator (MLE) of R. It may be mentioned here that related problems have been widely used in the statistical literature. The MLE of $P(Y < X)$, when X and Y have bivariate exponential distribution, has been considered by Awad *et al.* (1981).Church and Harris (1970), Downtown (1973), Govidarajulu (1967), Woodward and Kelley (1977) and Owen, Craswell and Hanson (1977) considered the estimation of $P(Y < X)$, when X and Y are normally distributed. Similar problem for the multivariate normal distribution has been considered by Gupta and Gupta (1990) . Kelley, Kelley and Schucany (1976), Sathe

and Shah (1981) considered the estimation of $P(Y < X)$ when X and Y are independent exponential random variables. Constantine and Karson (1986) considered the estimation of $P(Y < X)$, when X and Y are independent gamma random variables. Sathe and Dixit (2001) have been estimate of $P(Y < X)$ in the negative binomial distribution. Surles and Padgett (2001) considered the estimation of $P(Y < X)$, where X and Y are Burr Type X random variables. Baklizi and Dayyeh (2003) have done shrinkage estimation of $P(Y < X)$ in exponential case. Kundu and Gupta (2005) considered the estimation of $P(Y < X)$, where X and Y have generalized exponential distribution. Nasiri and Jabbari Nooghabi (2009) considered the estimation of $P(Y < X)$, where X and Y independent generalized exponential random variable in the presence of one outlier, when the scale parameter (λ) is known.

The rest of the paper is organized as follows; In section 2, we derive the MLE of R. Analysis of a real life data set has been presented in section 3 and finally we draw conclusion in section 4.

2. Maximum Likelihood estimator of R

Let $Y_1, Y_2, ..., Y_m$ be a random sample for Y with pdf

$$g(y, \alpha) = \alpha e^{-y}(1 - e^{-y})^{\alpha-1}, \qquad y > 0, \tag{7}$$

and $X_1, X_2, ..., X_n$ be random sample for X with pdf

$$f(x, \beta_1, \beta_2) = \frac{k}{n}\beta_1 e^{-x}(1 - e^{-x})^{\beta_1-1} + \frac{n-k}{n}\beta_2 e^{-x}(1 - e^{-x})^{\beta_2-1}, \qquad x > 0. \tag{8}$$

Then

$$\begin{aligned}
R &= P(Y < X) \\
&= \int_0^\infty \int_0^x g(y, \alpha) f(x, \beta_1, \beta_2,) dy dx \\
&= \int_0^\infty [\int_0^x \alpha e^{-y}(1 - e^{-y})^{\alpha-1} dy][\frac{k}{n}\beta_1 e^{-x}(1 - e^{-x})^{\beta_1-1} \\
&\quad + \frac{n-k}{n}\beta_2 e^{-x}(1 - e^{-x})^{\beta_2-1}]dx \\
&= \int_0^\infty \int_0^x [\frac{k}{n}\beta_1 e^{-x}(1 - e^{-x})^{\beta_1-1}][\alpha e^{-y}(1 - e^{-y})^{\alpha-1}] dy dx \\
&\quad + \int_0^\infty \int_0^x [\frac{n-k}{n}\beta_2 e^{-x}(1 - e^{-x})^{\beta_2-1}][\alpha e^{-y}(1 - e^{-y})^{\alpha-1}] dy dx \\
&= \int_0^\infty \frac{k}{n}\beta_1 e^{-x}(1 - e^{-x})^{\alpha+\beta_1-1} dx + \int_0^\infty \frac{n-k}{n}\beta_2 e^{-x}(1 - e^{-x})^{\alpha+\beta_2-1} dx \\
&= \frac{k}{n}\frac{\beta_1}{\alpha + \beta_1} + \frac{n-k}{n}\frac{\beta_2}{\alpha + \beta_2}.
\end{aligned} \tag{9}$$

Therefore, the MLE of R becomes

$$\hat{R} = \frac{k}{n}\frac{\hat{\beta}_1}{\hat{\alpha}+\hat{\beta}_1} + \frac{n-k}{n}\frac{\hat{\beta}_2}{\hat{\alpha}+\hat{\beta}_2}$$

For $k = 0$ (or in case of no outlier presence) the above equation obtained by by Kundu and Gupta (2005), and for $k = 1$ proposed by Nasiri and Jabbari Nooghabi (2009).

Now to compute the MLE of R, first we consider the joint distribution of $X_1, X_2, ..., X_n$with presence of k outliers in (5), so

$$\begin{aligned} &L(\alpha,\beta_1,\beta_2) \\ &\quad = g(y_1,y_2,...,y_m).f(x_1,x_2,...,x_n) \\ &\quad = \alpha^m e^{-\sum_{i=1}^m y_i}\prod_{i=1}^m (1-e^{-y_i})^{\alpha-1} \\ &\quad \times \frac{k!(n-k)}{n!}\,{\beta_1}^k \beta_2^{n-k} e^{-\sum_{i=1}^n x_i}\prod_{i=1}^n (1-e^{-x_i})^{\beta_2-1}\sum_{\underline{A}}\prod_{r=1}^k (1-e^{-x_{A_r}})^{\beta_1-\beta_2}. \end{aligned}$$

The Log-likelihood function of the observed sample is

$$\begin{aligned} \ln L(\alpha,\beta_1,\beta_2) &= m\ln(\alpha) - \sum_{i=1}^m y_i + (\alpha-1)\sum_{i=1}^m \ln(1-e^{-y_i}) \\ &\quad + C' + k\ln\beta_1 + (n-k)\ln\beta_2 - \sum_{i=1}^n x_i + (\beta_2-1)\sum_{i=1}^n \ln(1-e^{-x_i}) \\ &\quad + \ln\left(\sum_{\underline{A}}\prod_{r=1}^k (1-e^{-x_{A_r}})^{\beta_1-\beta_2}\right). \end{aligned} \tag{10}$$

where $C' = \ln\frac{k!(n-k)}{n!}$ and $\sum_{\underline{A}} = \sum_{A_1=1}^{n-k+1}\sum_{A_2=A_1+1}^{n-k+2}\cdots\sum_{A_k=A_{k-1}+1}^{n}$.

The MLE's of α, β_1 and β_2 say $\hat{\alpha}, \hat{\beta}_1$ and $\hat{\beta}_2$, respectively, which is obtained as the solutions of

$$\frac{\partial \ln L}{\partial\alpha} = \frac{m}{\alpha} + \sum_{i=1}^m \ln(1-e^{-y_i}) = 0$$

or

$$\frac{m}{\alpha} = -\sum_{i=1}^m \ln(1-e^{-y_i})$$

Hence

$$\hat{\alpha} = \frac{-m}{\sum_{i=1}^m \ln(1-e^{-y_i})}, \tag{11}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \beta_1} &= \frac{k}{\beta_1} + \frac{\frac{\partial}{\partial \beta_1}\sum_{\underline{A}}\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}}{\sum_{\underline{A}}\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \\ &= \frac{k}{\beta_1} + \frac{\sum_{\underline{A}}\left(\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}\sum_{r=1}^{k}\ln(1-e^{-x_{A_r}})\right)}{\sum_{\underline{A}}\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \\ &= 0, \end{aligned} \tag{12}$$

and

$$\begin{aligned}\frac{\partial \ln L}{\partial \beta_2} &= \frac{n-k}{\beta_2} + \sum_{i=1}^{n}\ln(1-e^{-x_i}) + \frac{\frac{\partial}{\partial \beta_2}\sum_{\underline{A}}\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}}{\sum_{\underline{A}}\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \\ &= \frac{n-k}{\beta_2} + \sum_{i=1}^{n}\ln(1-e^{-x_i}) \\ &\quad - \frac{\sum_{\underline{A}}\left(\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}\sum_{r=1}^{k}\ln(1-e^{-x_{A_r}})\right)}{\sum_{\underline{A}}\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \\ &= 0. \end{aligned} \tag{13}$$

There is no closed-form solution to this system of equations, so we will solve for $\hat{\beta}_1$ and $\hat{\beta}_2$ iteratively, using the Newton-Raphson method, a tangent method for root finding. In our case we will estimate $\theta = (\beta_1, \beta_2)'$ iteratively:

$$\hat{\theta}_{i+1} = \hat{\theta}_i - G^{-1}\mathbf{g} \ , \tag{14}$$

where g is the vector of normal equations for which we want

$$\mathbf{g} = [\ g_1 \quad g_2\ ] \tag{15}$$

with

$$g_1 = \frac{k}{\beta_1} + \frac{\sum_{\underline{A}}\left(\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}\sum_{r=1}^{k}\ln(1-e^{-x_{A_r}})\right)}{\sum_{\underline{A}}\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \ , \tag{16}$$

$$g_2 = \frac{n-k}{\beta_2} + \sum_{i=1}^{n}\ln(1-e^{-x_i}) - \frac{\sum_{\underline{A}}\left(\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}\sum_{r=1}^{k}\ln(1-e^{-x_{A_r}})\right)}{\sum_{\underline{A}}\prod_{r=1}^{k}(1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \ , \tag{17}$$

and G is the matrix of second derivatives

$$G = \begin{bmatrix} \frac{d\ g_1}{d\beta_1} & \frac{d\ g_1}{d\beta_2} \\ \\ \frac{d\ g_2}{d\beta_1} & \frac{d\ g_2}{d\beta_2} \end{bmatrix} \tag{18}$$

where

$$\frac{d\ g_1}{d\beta_1} = \frac{-k}{{\beta_1}^2} + \frac{\sum_{\underline{A}} \left(\prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2} \cdot \left(\sum_{r=1}^{k} \ln(1-e^{-x_{A_r}}) \right)^2 \right)}{\sum_{\underline{A}} \prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2}} + \left(\frac{\sum_{\underline{A}} \left(\prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2} \sum_{r=1}^{k} \ln(1-e^{-x_{A_r}}) \right)}{\sum_{\underline{A}} \prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \right)^2 \tag{19}$$

$$\frac{d\ g_1}{d\beta_2} = \frac{d\ g_2}{d\beta_1} = -\frac{\sum_{\underline{A}} \left(\prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2} \cdot \left(\sum_{r=1}^{k} \ln(1-e^{-x_{A_r}}) \right)^2 \right)}{\sum_{\underline{A}} \prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2}} + \left(\frac{\sum_{\underline{A}} \left(\prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2} \sum_{r=1}^{k} \ln(1-e^{-x_{A_r}}) \right)}{\sum_{\underline{A}} \prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \right)^2 \tag{20}$$

$$\frac{d\ g_2}{d\beta_2} = \frac{k-n}{{\beta_2}^2} + \frac{\sum_{\underline{A}} \left(\prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2} \cdot \left(\sum_{r=1}^{k} \ln(1-e^{-x_{A_r}}) \right)^2 \right)}{\sum_{\underline{A}} \prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2}} - \left(\frac{\sum_{\underline{A}} \left(\prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2} \sum_{r=1}^{k} \ln(1-e^{-x_{A_r}}) \right)}{\sum_{\underline{A}} \prod_{r=1}^{k} (1-e^{-x_{A_r}})^{\beta_1-\beta_2}} \right)^2 \tag{21}$$

The Newton-Raphson algorithm converges, as our estimates of β_1 and β_2 change by less than a tolerated amount with each successive iteration, to $\hat{\beta}_1$ and $\hat{\beta}_2$.

For $\beta_1 = \beta_2 = \beta$ in case of no outlier presence ($k = 0$), $\hat{\alpha}$ and $\hat{\beta}$ can be obtain as

$$\hat{\alpha} = \frac{-m}{\sum_{i=1}^{m} \ln(1-e^{-y_i})} \qquad and \qquad \hat{\beta} = \frac{-n}{\sum_{i=1}^{n} \ln(1-e^{-x_i})}. \tag{22}$$

For $\beta_1 = \beta_2$ (or $k = 0$), these equations (in (22)) were proposed by Kundu and Gupta (2005). For $k = 1$, the equations of (12) and (13) proposed by Nasiri and Jabbari Nooghabi (2009).

3. Numerical experiments and discussions

In order to have some idea about Bias and Mean Square Error (MSE) of MLE, we perform sampling experiments using a MATLAB. To have inference about R, we consider the following small sample size;

$$(m,n) = \\ (15,15),(20,20),(25,25),(15,20),(20,15),(15,25),(25,15),(20,25),(25,20).$$

Here, we take $\alpha = 1.50$, $\beta_1 = 2.75$ and $\beta_2 = 2.5$, for $k = 1, 2$. As we know, the generated sample size n from $f(x, \beta_1, \beta_2)$, $(n-k)$ sample generated from the equation (3) and k sample generated from the equation (4). All the results are base on 1000 replications. Here we present a complete analysis of a simulated data. The data has been generated using $m = n = 20$, $\alpha = 1.5, \beta_1 = 2.75$, $\beta_2 = 2.5$ and $k = 3$, thus $R = 0.6283$.

The data has been presented below. The Y values are,

0.4923	1.3188	1.6566	1.8956	0.7317	0.8856	0.2156
0.1871	1.1989	0.9144	0.8894	2.1774	0.3848	1.8992
1.3030	0.9214	0.3402	1.0771	1.8312	0.1487	

and the corresponding X values are ,

0.7019	1.9499	0.2770	1.7058	2.4813	2.5566	0.8180
1.5875	0.6742	0.8825	1.9649	1.3268	0.6253	3.5434
0.9046	0.5068	2.4653	1.6606	2.3153	3.0868	

Now, we obtain the MLE of $\hat{\alpha} = 1.4611, \hat{\beta}_1 = 2.6378$ and $\hat{\beta}_2 = 2.5261$. Therefore, $\hat{R} = 0.7301$.

4. Conclusions

In this paper, we have addressed the problem of estimating $P(Y < X)$ for the Generalized Exponential distribution with presence of k outliers when scale parameter is know. The results are given in Tables 1 and 2. It is observed that the maximum likelihood estimator R work quit well. We report the average estimates and the MSEs based on 1000 replications. In this case as expected when $m = n$ and m, n increase then the average biases and the MSEs decrease . For fixed m as n increases the MSEs decrease and also for fixed n as m increases the MESs decreases.

Table 1

$k = 1 \quad , \quad \alpha = 1.5 \quad , \quad \beta_1 = 2.75 \quad , \quad \beta_2 = 2.5$

(n,m)	Bias $\hat{\alpha}$	MSE $\hat{\alpha}$	Bias $\hat{\beta_1}$	MSE $\hat{\beta_1}$	Bias $\hat{\beta_2}$	MSE $\hat{\beta_2}$	Bias $\hat{R}$	MSE $\hat{R}$
(15,15)	0.1421	0.2527	-0.2421	383.24	0.2526	0.9559	-0.3097	0.1838
(20,20)	0.0808	0.1555	-1.2602	14.254	0.1824	0.6091	-0.3124	0.1883
(25,25)	0.0660	0.1135	0.8531	1509.6	0.1417	0.4601	-0.3013	0.1812
(15,20)	0.0751	0.1467	4.3815	19357	0.2523	0.9969	-0.2966	0.1768
(20,15)	0.1182	0.2183	-1.2712	20.414	0.1503	0.5981	-0.3026	0.1805
(15,25)	0.0633	0.1069	-0.8652	33.742	0.2195	0.8879	-0.2908	0.1720
(15,25)	0.1064	0.2132	-0.2088	628.51	0.1138	0.4227	-0.2973	0.1792
(20,25)	0.6214	0.1177	1.9963	86705	0.1859	0.6329	-0.3086	0.1855
(25,20)	0.0818	0.1455	2.0314	5162.2	0.2115	0.4016	-0.3049	0.1848

Table 2

$k = 2 \quad , \quad \alpha = 1.5 \quad , \quad \beta_1 = 2.75 \quad , \quad \beta_2 = 2.5$

(n,m)	Bias $\hat{\alpha}$	MSE $\hat{\alpha}$	Bias $\hat{\beta_1}$	MSE $\hat{\beta_1}$	Bias $\hat{\beta_2}$	MSE $\hat{\beta_2}$	Bias $\hat{R}$	MSE $\hat{R}$
(15,15)	-0.0126	0.0064	-0.0984	0.0669	-0.1962	0.0681	-0.0070	0.0073
(20,20)	-0.0103	0.0044	-0.0481	0.0745	0.0048	0.0413	-0.0020	0.0056
(25,25)	-0.0069	0.0033	0.1930	0.0579	0.0121	0.0174	-0.0025	0.0041
(15,20)	-0.0154	0.0052	0.2274	0.0635	0.0648	0.0098	-0.0058	0.0061
(20,15)	-0.0172	0.0064	0.2878	0.0933	0.2524	0.0812	-0.0065	0.0063
(15,25)	-0.0099	0.0043	0.2812	0.0832	0.1053	0.0288	0.0020	0.0055
(15,25)	-0.0176	0.0054	0.0463	0.0729	0.0397	0.0416	-0.0085	0.0056
(20,25)	-0.0096	0.0040	-0.0990	0.0238	0.0861	0.0994	-0.0026	0.0050
(25,20)	-0.0088	0.0043	-0.0163	0.0469	-0.0926	0.0580	-0.0018	0.0050

References

1. Awad, A. M., Azzam, M. M. and hamadan, M. A. Some inference results in $P(Y < X)$ ine bivariate expoential model. Communications in statistics - *Theroy and Methods, 1981, Vol. 10, 2515-2524.*
2. Baklizi, A. and Dayyh, W. A. Shrinkage estimation of $P(Y < X)$ in the exponential case. *Comm. in statistics, Theory and Methods 2003, 32(1) : 31-42.*
3. Church, J.D. and Harris, B. The estimation of reliablity from stress strength relationships. *Technometrics, 1970, Vol. 12, 49-54.*
4. Constantine, K. and Karson, M. The estimation of $P(Y < X)$ in gamma case. *Communications in statistics - Computations and simulations, 1986, Vol. 15, 365-388.*
5. Downtown, F. The estimation of $P(Y < X)$ in normal case. *Technometrics, 1973, Vol. 15, 551-558.*
6. Dixit, U. J. (1989). Estimation of parameters of the Gamma distribution in the presence of outliers. *Comm. in statistics, Theory and methods 18, 3071-3085*
7. Dixit, U. J. and Nasiri, P. F. (2001). Estimation of parameters of the exponential distribution in the presence of outliers generated from uniform distribution, *Metron 49(3-4) : 187-198.*
8. Gupta, R.D. and Gupta, R.C. Estimation of $P(aX > bY)$ in the multivarite normal case. *Statistics. 1990, Vol. 1, 91-97.*
9. Gupta, R.D. and Kundu, D. Generalied exponential distributions. *Australian and Newzealand Journal of Stastistics, 1999, Vol. 41, 173-188.*
10. Gupta, R.D. and Kundu, D. Generalied exponential distributions; Different Method of Esimations. *Journal of Stastistics Computation and simulation, 2001a, Vol. 69, 315-338.*
11. Gupta, R.D. and Kundu, D. Generalied exponential distributions; An alternative to gamma or Weibull distribution. *Biometrical Journal, 2001b, Vol. 43, 117-130.*
12. Gupta, R.D. and Kundu, D. Generalied exponential distributions; Statistical Inferences. *Journal of Statistical Theory and Applications, 2002, Vol. 1, 101-118.*
13. Gupta, R.D. and Kundu, D. Closeness between the gamma and generalized exponentioal distributions. *Communications in Statistcs - Theory and Methods, 2003a, Vol. 32, 705-722.*
14. Gupta, R.D. and Kundu, D. Discriminating between the Weibull and Generalied exponential distributions. *Computational Statistics and Data Analysis, 2003b, Vol. 43, 179-196.*
15. Gupta, R.D. and Kundu, D. Discriminating between gamma and the Generalied exponential distributions. *Journal of Statistical Computation and Simulation, 2004, Vol. 74, 107-122.*
16. Kelley, G.D., Kelley, J.A. and Schucany, W.R. Efficient estimation of $P(Y < X)$ in expoential case. *Technometrics, 1976, Vol. 15, 359-360.*
17. Kundu, D. and Gupta R. D. Estimation of $P[Y < X]$ for generalied exponentioal distribution. *Metrika, 2005, Vol. 61, No. 3 pp. 291-308.*
18. Mudholkar, G.S. and Srivastava, D.K. Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEEE Transactions on Reliability, Vol. 42, 1993, 299-302.*
19. Mudholkar, G.S. and Hutson, A.D. The exponentiated Weibull family : some prop-

erties and a flood data applications. *Communications in Statistics - Theory and Methods, 1996, Vol. 25, 3059-3083.*

20. Mudholkar, G.S. and Srivastava, D.K. and Freimer, M. The exponentiated Weibull family: a reanalysis of the bus-motor failure data. *Technometrics, 1995, Vol. 37, 436-445.*
21. Nasiri, P. and Jabbari Nooghabi, M. Estimation of $P[Y < X]$ for generalized exponential distribution in the presence of outlier. *Mashhad Research Journal of Mathematical Sciences, 2009, Vol. 2, 71-80.*
22. Owen, D.B., Craswell, K.J. and Hanson, D.L. Non-parametric upper confidence bounds for $P[Y < X]$ and confidence limits for $P(Y < X)$ when X and Y are normal. *Journal of yhe American Statistical Association, 1977, Vol. 59, 906-924.*
23. Ragab, M.Z. Inference for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference, 2002, Vol. 104, 339-350.*
24. Ragab, M.Z. and Ahsanullah, M. Estimation of the location and parameters of the generalized exponential distribution based on order statistics. *Journal of Statistical Computation and Simulation, 2001, Vol. 69, 109-124.*
25. Sathe, Y. S. and Dixit, U. J. (2001). Estimation of $(X \leq Y)$ in negative binomial distribution. *Journal of statistical planning and Inference 03, 83-92.*
26. Sathe, Y.S. and Shah, S.P. On estimation $P(Y < X)$ for the exponential distribution. *Communications in Statistics-Theory and Methods, 1981, Vol. 10, 39-47.*
27. Surles, J.G. and Padgett, W.J. Inference for reliablitiy and stress-strength for a scaled Burr-type X distribution. *Lifetime Data Analysis, 2001, Vol. 7, 187-200.*
28. Woodward, W.A. and Kelley, G.D. Minimum variance unbiased estimation of $P[Y < X]$ in the normal case. *Technometrics, 1977, Vol. 19, 95-98* .
29. Zheng, G. On the Fisher information matrix in type-II censored data from the exponentiaed exponetial family. *Biometrical Journal, 2002, Vol. 44, 353-357.*

Structure Of Generalized Bivariate Lomax Distribution Based On Dependence And Information Measures

V. Ranjbar, M. Amini, A. Bozorgnia and G.R. Mohtashami Borzadaran

In this paper, the dependence structure of the generalized bivariate Lomax family studied. First, some concepts of dependence and some association measures like tail dependence and extremal dependence obtained. In addition, some local dependence functions derived and compared via a numerical study. finally, entropy, mutual information and quadratic mutual information measures for this family studied and discussed about them.

Keywords: Copula function, Dependence measures, Generalized bivariate Lomax distribution, Entropy, Mutual information, Quadratic mutual information, Local dependence function, Heavy tailed distributions.

1. Introduction and Preliminaries

In recent years, a number of studies in economics, finance, networking and some other sciences like these, have focused on dependence measuring, modeling, information measuring and heavy tailed distributions. In addition, the bivariate measures of dependence and the copula based approaches to dependence modeling are two interrelated parts of the study of dependence structure of bivariate distributions in mathematical statistics and probability theory. Many authors studied dependence structures of some bivariate distributions such as: Shaked (1977) has presented some concepts of dependence for bivariate distributions, Schweizer and Wolff (1981) obtained nonparametric measures of dependence for random variables. Apparently, some authors wrote useful papers in the field of dependence via computing well known dependence measures for some bivariate distributions, for example, Bairamove and Kotz (2002) studied the dependence structure of Farlie-Gumbel-Morgenstern distributions and their extensions, Nadarajah et al. (2003) determined the local dependence function for extreme value distributions in view of Kotz and Nadarajah's (2003) Local dependence function. A new measure of linear local dependence has been obtained by Bairamov et. al. (2003), Sankaran and Gupta (2004) studied the properties of the local dependence function, introduced by Holland and Wang (1987). Moreover, they presented characterizations for bivariate Lomax distribution, bivariate Dirichlet distribution, and bivariate normal distribution using local function and regression functions and Asadian et al.(2008) investigated aspects of dependence in Lomax family. Cuadras and Auge (1981) introduced the family of bivariate distributions and investigated dependence structure of it, Cuadras (2006) derived dependence measures, Kendall's tau

and Spearman's rho in Cuadras-Auge family and expanded this bivariate distribution in terms of Fréchet-Hoeffding lower and upper bounds, Ruiz-Rivas and Cuadras (1988) studied some geometrical, probabilistic and statistical properties of the Cuadras-Ague family and Bolbolian et al.(2008) studied dependence structure of Cuadras-Auge family via some measure of association and tail dependence coefficients. Genest (1987) considered the dependence structure of Frank's family of bivariate distribution, also suggested three nonparametric estimator of the association parameter and compared their small-sample behavior of the Maximum likelihood estimator for this family.
The other way to determine the measure of dependence between two random variables is using the information theory. Some measures such as entropy, mutual information and quadratic mutual information play important role in dependence measuring of bivariate distributions and some papers have written in this case. For example, Joe (1989) has presented the relative entropy measures of multivariate dependence. Bell (1962) has used mutual information as a measure of dependence. Xu and Principe (1998) discussed a novel algorithm to train nonlinear mappers with information theoretic criteria (entropy and mutual information) directly from a training set.
Also, the heavy tailed random variables their asymptotic behaviors and applications have been extensively investigated in half past century by many authors. The heavy tailed random variables have very considerable role in some sciences like finance, insurance and economics and study of the structures of the distributions of these random variables is one of the intrusting topics for statisticians. In particular, the study of the dependence structure of tail of distribution of random variables is considerable. Frahm et al. (2005) derived some properties of estimating the tail dependence coefficient. Caillault and Guegan (2005) have introduced non-parametric estimators for upper and lower tail dependence and confidence intervals are obtained with a bootstrap method, Dobric and Schmid (2005) estimated the lower tail dependence in bivariate copulas by nonparametric approach. Many other authors presented some paper in this case such as Juri and Wüthrich (2003), Charpentier and Segers (2006).
In view of these themes, we want to study the dependence structure of a family of bivariate distributions which contains the distributions with heavy tailed marginal distributions via some dependence structure and information measures.
Let (X, Y) be a random vector with the following survival distribution function,

$$\overline{F}(x, y) = (1 + a_1 x^{a_2} + b_1 y^{b_2})^{-p}, \quad x, y \geq 0, \tag{1}$$

where a_1, a_2, b_1, b_2 and p are positive real numbers. This family of bivariate distributions is extended version of bivariate Lomax distribution, that is called "generalized bivariate Lomax (GBL)". This class contain two main categories of the bivariate distributions such as:

- Bivariate Lomax distributions for $a_2 = b_2 = 1$, which is widely used in reliability theory(see for details, Nayak, 1987, Nadarajah, 2005 and Barlow and Proschan 1981).
- Bivariate Burr distributions for $a_1 = b_1 = 1$, that belongs to the class of heavy tailed distribution. It has many applications in finance, insurance and networking (see, Resnik, 2006).

The purpose of this paper is to examine the dependence structure of GBL family. We investigate dependence structure of this family via computing measures of information and dependency. In some cases, we have used copula function, instead of distribution function, since computations are simple and the copula is independent from marginal distributions. .
The rest of this paper consists of seven sections. After listing some concepts of dependence and surveying them for GBL family in section 2, we derive some association measures, tail dependence coefficients, extremal dependence coefficients in Section 3, and also we compare these coefficients with a numerical approach in this section. We discuss the behavior of the local dependence via Bairamov and Kotz local dependence function and Clayton-Oakes association measure in GBL family and draw graphs of these local dependence measures in section 4. In section 5, we compute three information measures and evaluate them for this family. Finally, after writing a short conclusion in section 6, we present the proofs of the theorems and lemmas in section 7.

2. Some Concepts of Dependence

Let (X, Y) be a random vector with joint density function $f(x, y)$, distribution function $F(x, y)$ and marginals $F_1(x)$ and $F_2(x)$. Then the following quantities are defined:

1. (Karlin, 1968) The real function $h(x, y)$ is totally positive of order two (TP_2) if $h(x, y) \geq 0$,
$$h(x, y)h(x', y') \geq h(x, y')h(x', y), \quad \forall \ x < x', y < y'.$$
2. The random vector (X, Y) is said to be positive likelihood ratio dependence ($PLRD(X, Y)$) if $f(x, y)$ is TP_2.
3. The random vector (X, Y) or its distribution function is said to be right corner set increasing ($RCSI(X, Y)$) if $P(X > x, Y > y|X > x', Y > y')$ is increasing in x' and y' for all x and y.
4. The random variable X is said to be stochastically increasing in Y ($SI(X|Y)$) if $P(X > x|Y = y)$ is increasing in y for all x.
5. The random variable X is said to be right tail increasing in Y ($RTI(X|Y)$) if $P(X > x|Y > y)$ is non-decreasing in y for all x.
6. The random variable X is said to be left tail decreasing in Y ($LTD(X|Y)$) if $P(X \leq x|Y \leq y)$ is non-increasing in y for all x.
7. The random variables X and Y are said to be positively quadrant dependent ($PQD(X, Y)$) if,
$$P(X > x, Y > y) \geq P(X > x)P(Y > y), \quad \forall x, y.$$
8. The bivariate failure (or hazard) rate of a random vector (X, Y) is defined as follows,
$$r(x, y) = \frac{f(x, y)}{\bar{F}(x, y)}. \tag{2}$$

9. The copula function $C(u,v)$ is a bivariate distribution function with uniform marginals on $[0,1]$, such that

$$F(x,y) = C_F(F_1(x), F_2(y))$$

By Sklar's Theorem (Sklar, 1959), this copula exists and is unique if F_1 and F_2 are continuous. Thus we can construct bivariate distributions $F(x,y) = C_F(F_1(x), F_2(y))$ with given univariate marginals F_1 and F_2 by using copula C_F,(Nelsen, 2006).

We have the following properties for copula functions:

- (Nelsen, 2006) Let $F(x,y)$ be a joint distribution function with marginals $F_1(x)$ and $F_2(y)$, then
  (i) The copula C_F is given by

$$C_F(u,v) = F(F_1^{-1}(u), F_2^{-1}(v)), \quad \forall u,v \in [0,1],$$

  where, F_1^{-1} and F_2^{-1} are quasi-inverses of F_1 and F_2 respectively.
  (ii) The partial derivatives $\partial C_F(u,v)/\partial u$ and $\partial C_F(u,v)/\partial v$ exist and $c(u,v) = \partial^2 C_F(u,v)/\partial u \partial v$ is density function of $C_F(u,v)$.
- The random vector (X,Y) is $PLRD$ if and only if

$$\frac{c(u',v)}{c(u,v)} \quad \textit{is incraesing in } v \textit{ for all } u < u'. \tag{3}$$

Remark 2.1. (i)- Let (X,Y) be a random vector with GBL distribution function, then, using relation (1), it is easy to see that,

$$C(u,v) = \left[(1-u)^{-\frac{1}{p}} + (1-v)^{-\frac{1}{p}} - 1\right]^{-p} + u + v - 1, \quad p > 0. \tag{4}$$

(ii)- We observe that the copula function of GBL family is independent of parameters $a_i, b_i, i = 1,2$, furthermore, it is equal to copula of bivariate Lomax family that has been studied by Asadian et al.(2008).

Proposition 2.1. *Let (X,Y) be a random vector with GBL distribution function, then (X,Y) is PLRD and consequently is PQD.*

Proof. Using relation (4), we get,

$$\begin{aligned} c(u,v) &= \frac{\partial^2 C(u,v)}{\partial v \partial u} \\ &= \frac{p+1}{p}(1-v)^{-\frac{1}{p}-1}(1-u)^{-\frac{1}{p}-1}\left[(1-u)^{-\frac{1}{p}} + (1-v)^{-\frac{1}{p}} - 1\right]^{-p-2}, \end{aligned}$$

So, we have

$$\frac{c(u',v)}{c(u,v)} = \left(\frac{1-u}{1-u'}\right)^{\frac{p+1}{p}} \left[1 - \frac{(1-u')^{\frac{-1}{p}} - (1-u)^{\frac{-1}{p}}}{(1-u')^{\frac{-1}{p}} + (1-v)^{\frac{-1}{p}} - 1}\right]^{p+2}, \quad \forall u < u', 0 < v < 1,$$

which is increasing function in v and this completes the proof. □

Proposition 2.2. *Let (X, Y) be a random vector with GBL function, then*

(i). X is stochastically increasing in Y,$(SI(X|Y))$ and Y is stochastically increasing in X $(SI(Y|X))$.

(ii). X is right tail increasing in Y,$(RTI(X|Y))$ and Y is right tail increasing in X $(RTI(Y|X))$.

(iii). X is left tail decreasing in Y,$(LTD(X|Y))$ and Y is left tail decreasing in X $(LTD(Y|X))$.

Proof. Using relation (1) we have

$$P(Y > y|X = x) = \Big(1 + \frac{b_1 y^{b_2}}{1 + a_1 x^{a_2}}\Big)^{-p-1},$$

$$P(Y > y|X > x) = \Big(1 + \frac{b_1 y^{b_2}}{1 + a_1 x^{a_2}}\Big)^{-p},$$

$$P(Y \leq y|X \leq x) = 1 + \frac{(1 + a_1 x^{a_2} + b_1 y^{b_2})^{-p} - (1 + b_1 y^{b_2})^{-p}}{1 - (1 + a_1 x^{a_2})^{-p}}.$$

It is easy to see that the first two relation are increasing function of x and the last one is decreasing, which completes the proof. □

Remark 2.2. The random vector (X, Y) is $RCSI$ if $\bar{F}(x, y)$ is TP_2. (Nelsen 2006, Theorem 5.2.15)

Corollary 2.1. *Let (X, Y) be a random vector with GBL distribution function. It is easy to see that $\bar{F}(x, y)$ is TP_2 and then (X, Y) is $RCSI$.*

3. Some measures of association

In this section, we compute measures of association, tail dependence coefficients and extremal dependence coefficients for GBL family.

3.1. *Kendall's τ and Spearman's ρ_s*

The Spearman's ρ_s is connected with PQD concept and formulated by copula function C as follows:

$$\rho_s = 12 \int_0^1 \int_0^1 [C(u, v) - uv] du dv = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3.$$

The Kendall's τ is connected with PLRD concepts and formulated by copula function C as:

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 = 1 - 4 \int_0^1 \int_0^1 \Big[\frac{\partial C}{\partial u} . \frac{\partial C}{\partial v}\Big] du dv.$$

These measures are obtained for bivariate Lomax family by Asadian et al.(2008), and since the copula of GBL family is equal to copula of bivariate Lomax family, so the following proposition is valid for GBL family.

Proposition 3.1. *Let (X, Y) be a random vector with GBL distribution function, then,*
(i). $\tau = \frac{1}{2p+1}$,
(ii). $\rho_s = \sum_{k=0}^{\infty} \frac{12p^2(k+1)}{2p+k} B(3p, k+1) - 3$,
where, $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ *is the Beta function and* $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$.

Remark 3.1. Let (X, Y) be a random vector with GBL distribution function, then, Proposition 2.3, Theorem 5.1 in Fredricks and Nelsen, (2007) and Hutchinson and Lai (1990)[Exercise, 5.38 in Nelsen, 2006] imply that

$$0 \leq \tau \leq \rho_s \leq \frac{3}{2}\tau.$$

3.2. *The Blomqvist medial coefficient*

This coefficient, also known as quadrant test of Blomqvist, evaluates the dependence at the center of a distribution (Blomqvist, 1950). If X and Y are independent, then in particular $C(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4}$. The coefficient of Blomqvist is defined as:

$$\beta = 4C(\frac{1}{2}, \frac{1}{2}) - 1.$$

Corollary 3.1. *Let (X, Y) be a random vector with GBL distribution function, then,* $\beta = 4(2^{(p+1)/p} - 1)^{-p} - 1$. *It is obvious that* $0 \leq \beta \leq 1$ *for all* $p > 0$.

3.3. *Schweizer-Wolff's index of dependence*

An index closely related to Spearmans ρ_s is the index σ_{XY} introduced by Schweizer and Wolff (1981). Instead of considering the difference $C(u, v) - uv$, they use its absolute value to define:

$$\sigma_{XY} = 12 \int_0^1 \int_0^1 |C(u, v) - uv| du dv,$$

which is a measure of the volume between the surfaces $C(u, v)$ and uv. We have the following implications

$$\sigma_{XY} = 0 \Rightarrow (X, Y) \textit{ independent},$$
$$\sigma_{XY} = 1 \Rightarrow X \textit{ is a monotone function of } Y.$$

Corollary 3.2. *Since in GBL family,* $C(u, v) \geq uv$, *we obtain,*

$$\sigma_{XY} = \rho_s.$$

3.4. *Gini's gamma coefficient*

The Gini's γ coefficient is defined as

$$\gamma_C = 2\int_0^1\int_0^1 (|u+v-1|-|u-v|)dC(u,v).$$

Another form of Gini's γ is given by

$$\gamma_C = 4\left\{\int_0^1 C(u,1-u)du - \int_0^1 [u-C(u,u)]du\right\}.$$

(For more details see, Nelsen, 2006).

Corollary 3.3. *Let (X,Y) be a random vector with GBL distribution function, then, for all $p>0$,*

$$\gamma_C = 4\left[\int_0^1 \left([(1-u)^{\frac{-1}{p}} + u^{\frac{-1}{p}} - 1]^{-p} + [2(1-u)^{\frac{-1}{p}} - 1]^{-p}\right) du\right] - 2.$$

Remark 3.2. Since there is not closed form for the Gini's gamma coefficient in GBL family, we study and analysis of it with numerical approach. Table 1 presents Gini's gamma coefficient and Spearman's ρ (or σ_{XY}) for GBL family for some values of p. It is easy to see that $\rho = \sigma_{XY} \geq \gamma_C$ for all values of p.

Table 1: γ_C and ρ_s for some values of p in GBL family.

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
γ_C	.877	.776	.692	.622	.565	.516	.475	.439	.409	.382
$\rho_s = \sigma_{XY}$	.952	.885	.811	.743	.683	.631	.585	.545	.510	.479
p	2	3	4	5	6	7	8	9	10	20
γ_C	.230	.164	.128	.104	.088	.076	.067	.060	.055	.027
$\rho_s = \sigma_{XY}$	.295	.212	.166	.136	.115	.099	.088	.079	.071	.037
p	30	40	50	60	70	80	90	100	...	1000
γ_C	.018	.014	.011	.009	.008	.007	.006	.005	...	.0006
$\rho_s = \sigma_{XY}$	.025	.018	.015	.012	.011	.009	.008	.007	...	.0007

3.5. *Tail dependence coefficients*

Let (X,Y) be a random vector with joint distribution function F and marginals F_1 and F_2, respectively. The quantity $\lambda_u = \lim_{t\to 1^-} P\left(F_1(X) > t | F_2(Y) > t\right)$ is called the upper tail dependence coefficient (UTDC) provided the limit exists. We say that (X,Y) has upper tail dependence if $\lambda_u > 0$ and upper tail independent if $\lambda_u = 0$. Similarly, we define the lower tail dependence coefficient (LTDC) by $\lambda_l = \lim_{t\to 0^+} P\left(F_1(X) \leq t | F_2(Y) \leq t\right)$. The upper tail dependence coefficient (or

lower tail dependence coefficient) can also be defined via the notion of copula. If C(u,v) is the copula of (X, Y), then

$$\lambda_u = \lim_{t\to 1^-} \frac{1-2u+C(u,u)}{1-u} \quad and \quad \lambda_l = \lim_{t\to 0^+} \frac{C(u,u)}{u}.$$

For more details, see Coles et. al. (2000).

Corollary 3.4. *Let (X, Y) be a random vector with GBL distribution function. Then $\lambda_u = 2^{-p}$ and $\lambda_l = 0$*

3.6. *Extremal dependence coefficients*

Extremal dependence coefficients were introduced by Frahm (2006) for studying the asymptotic dependence structure of the minimum and the maximum of a random vector. Let $(X_1, X_2, ..., X_n)$ be a random vector with joint distribution function $F(x_1, x_2, ..., x_n)$ and marginal distribution functions $F_1, ..., F_n$. Moreover, let $F_{\min} = \min\{F_1(X_1), ..., F_n(X_n)\}$ and $F_{\max} = \max\{F_1(X_1), ..., F_n(X_n)\}$.
The lower extremal dependence coefficient (LEDC) of $(X_1, X_2, ..., X_n)$ is defined as

$$E_l = \lim_{t\to 0^+} P(F_{max} \le t | F_{min} \le t),$$

where the upper extremal dependence coefficient (UEDC) of $(X_1, X_2, ..., X_n)$ is defined as

$$E_u = \lim_{t\to 1^-} P(F_{min} > t | F_{max} > t),$$

provided the corresponding limits exist.

Remark 3.3. By Proposition 1 in Frahm (2006), we can derive E_l and E_u via the quantities λ_l and λ_u as follows

$$E_l = \frac{\lambda_l}{2-\lambda_l} \quad and \quad E_u = \frac{\lambda_u}{2-\lambda_u}.$$

Therefore if $F(x, y)$ belongs to GBL family, then the Corollary 3.4 implies that

$$E_l = 0, \quad and \quad E_u = \frac{1}{2^{p+1}+1} > 0.$$

This means that (X, Y) has UED but not LED.

Remark 3.4. Since we have closed forms expression for τ, β, λ_u and E_u for the GBL family, it is possible to compare analytically these measures. Figure 1 contains graphs of these functions with compare them for some value of p.

i. If $0 < p < 1$, then $E_u < \tau < \beta < \lambda_u$.
ii. If $1 \le p < 2.6598$, then $E_u < \beta \le \tau < \lambda_u$.
iii. If $2.6598 < p < 2.7211$, then $E_u < \beta < \lambda_u < \tau$.
iv. If $p > 2.7211$, then $E_u < \lambda_u < \beta < \tau$.

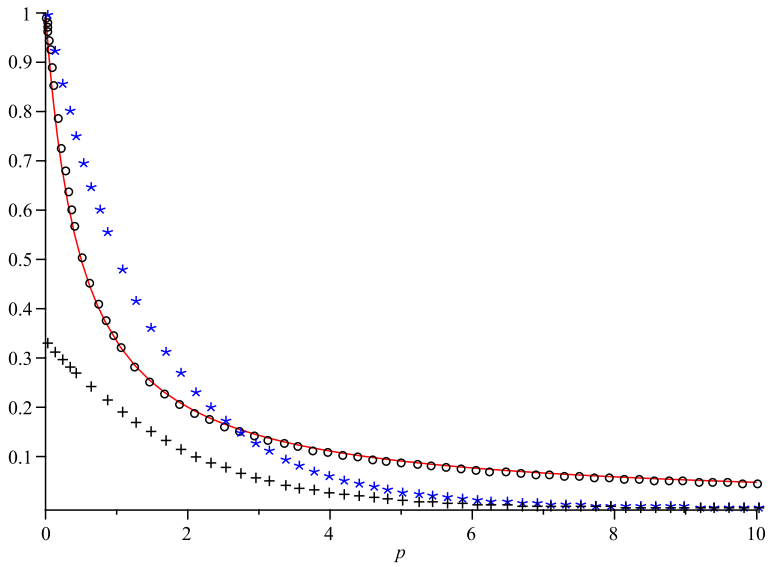

Figure 1: Comparing the dependence coefficients for some values of p, where λ_u shows by " $*$ ", E_u by " $\times$ ", β by " $\circ$ " and τ by line.

4. Local Dependence

In this section, we compute the Clayton-Oakes association measure (denoted by $\Theta(x,y)$) and local dependence function due to Kotz and Nadarajah (2003) (denoted by $H(x,y)$) in GBL family. Furthermore, we investigate the behavior of these measures drawing their graphs.

4.1. *Clayton-Oakes Association Measure*

Clayton (1978) and Oakes (1989) defined the following association measure.

$$\theta(x,y) = \frac{\bar{F}(x,y)D_{12}\bar{F}(x,y)}{D_1\bar{F}(x,y)D_2\bar{F}(x,y)},$$

where,

$$D_{12}\bar{F}(x,y) = \frac{\partial^2\bar{F}(x,y)}{\partial x\partial y};\ D_1\bar{F}(x,y) = \frac{\partial\bar{F}(x,y)}{\partial x};\ D_2\bar{F}(x,y) = \frac{\partial\bar{F}(x,y)}{\partial y}.$$

If $\theta(x,y) > 1$, we say X and Y are positive dependent and negative dependent if the opposite relation occurs. Consequently if $\theta(x,y) = 1$, X and Y being independent. Gupta (2003) proved that $\theta(x,y) = r(x|Y = y)/r(x|Y > y)$, where $r(x|Y = y) = f(x|y)/\bar{F}(x|y)$ and $r(x|Y > y)$ is the hazard rate of the conditional distribution of X given $Y > y$. Using this result we obtain the following Proposition.

Proposition 4.1. *Let (X, Y) be a random vector with GBL function, then*

$$\theta(x, y) = \frac{p+1}{p}.$$

Proof. By Gupta (2003) for every real value of x and y, we have

$$\begin{aligned} r(x|Y = y) &= -\frac{D_{12}\bar{F}(x, y)}{D_2\bar{F}(x, y)} \\ &= a_1a_2(p+1)x^{a_2-1}(1 + a_1x_{a_2} + b_1y^{b_2})^{-1}, \end{aligned} \tag{5}$$

and

$$\begin{aligned} r(x|Y > y) &= -\frac{D_1\bar{F}(x, y)}{\bar{F}(x, y)} \\ &= a_1a_2px^{a_2-1}(1 + a_1x_{a_2} + b_1y^{b_2})^{-1}. \end{aligned} \tag{6}$$

So we have

$$\theta(x, y) = \frac{r(x|Y = y)}{r(x|Y > y)} = \frac{p+1}{p}.$$

This completes the proof. □

4.2. *Local dependence function of Kotz and Nadarajah*

Kotz and Nadarajah (2003) have introduced a local dependence function (denoted by $H(x, y)$), which provides a local point of view on dependence at a point (x, y) and defined

$$H(x, y) = \frac{E\{(X - E(X|Y = y))(Y - E(Y|X = x))\}}{\sqrt{E(X - E(X|Y = y))^2}\sqrt{E(Y - E(Y|X = x))^2}},$$

which is obtained from the expression of the Pearson correlation coefficient by replacing mathematical expectations $E(X)$ and $E(Y)$ by conditional expectations $E(X|Y = y)$ and $E(Y|X = x)$, respectively. So, we have,

$$H(x, y) = \frac{\rho + \phi_X(y)\phi_Y(x)}{\sqrt{1 + \phi_X^2(y)}\sqrt{1 + \phi_Y^2(x)}}, \tag{7}$$

where,

$$\rho = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}};\ \phi_X(y) = \frac{E(X) - E(X|Y)}{\sqrt{var(X)}};\ \phi_Y(x) = \frac{E(Y) - E(Y|X)}{\sqrt{var(Y)}}.$$

Proposition 4.2. *Let (X, Y) be a random vector with GBL distribution function, then*

$$\phi_X(y) = \frac{B(p - \frac{1}{a_2}, 1 + \frac{1}{a_2})\left[p - (p - \frac{1}{a_2})(1 + b_1y^{b_2})^{\frac{1}{a_2}}\right]}{\sqrt{p}\left[B(p - \frac{2}{a_2}, 1 + \frac{2}{a_2}) - pB^2(p - \frac{1}{a_2}, 1 + \frac{1}{a_2})\right]^{\frac{1}{2}}},$$

$$\phi_Y(x) = \frac{B(p-\frac{1}{b_2}, 1+\frac{1}{b_2})\left[p-(p-\frac{1}{b_2})(1+a_1x^{a_2})^{\frac{1}{b_2}}\right]}{\sqrt{p}\left[B(p-\frac{2}{b_2}, 1+\frac{2}{b_2}) - pB^2(p-\frac{1}{b_2}, 1+\frac{1}{b_2})\right]^{\frac{1}{2}}},$$

and

$$\rho = \frac{\left[(p+1)B(p-\frac{1}{a_2}-\frac{1}{b_2}, 1+\frac{1}{a_2}, 1+\frac{1}{b_2}) - pB(p-\frac{1}{a_2}, 1+\frac{1}{a_2})B(p-\frac{1}{b_2}, 1+\frac{1}{b_2})\right]}{\left[\left(B(p-\frac{2}{a_2}, 1+\frac{2}{a_2}) - pB^2(p-\frac{1}{a_2}, 1+\frac{1}{a_2})\right)\left(B(p-\frac{2}{b_2}, 1+\frac{2}{b_2}) - pB^2(p-\frac{1}{b_2}, 1+\frac{1}{b_2})\right)\right]^{\frac{1}{2}}}.$$

By substituting theses relations in (7) the Kotz and Nadarajah dependence coefficient can be obtained, on noting that $B(a,b,c) = \Gamma(a)\Gamma(b)\Gamma(c)/\Gamma(a+b+c)$.

Corollary 4.1. *In particular if* $a_1 = a_2 = b_1 = b_2 = 1$*, the Katz and Nadarajah coefficient for GBL family is as follows:*

$$H(x,y) = \frac{p^3 + (p-2)^2[x(1-p)+1][y(1-p)+1]}{\sqrt{[p^4 + (p-2)^2(x(1-p)+1)^2][p^4 + (p-2)^2(y(1-p)+1)^2]}}.$$

As we see, this coefficient is a decreasing function of p and is tends to infinity when p tends to infinity. Figure 2 shows the behavior of $H(x,y)$ for values of $p = 1, 2, 4, 6, 10, 25$, under the conditions of Corollary 4.4.

Remark 4.1. We compute the saddle point (x^*, y^*) such that $H(x^*, y^*) = \rho$ when $a_1 = a_2 = b_1 = b_2 = 1$ in GBL family. By solving the equation $\phi_X(y^*) = \phi_Y(x^*) = 0$ analytically, we obtain

$$(x^*, y^*) = \left(\frac{1}{p-1}, \frac{1}{p-1}\right)$$

As we see, the Kotz and Nadarajah's local dependence are equal to Pearson's correlation coefficient ρ when X and Y are equal to their expectation. It is easy to see that in this case $H(x^*, y^*) = \rho = \frac{1}{p}$.

In order to compar Pearson's ρ, Clayton-Oakes association measure and local dependence function of Kotz and Nadarajah, for some value of p and (x^*, y^*), Table 2 presents ρ, and $\Theta(x^*, y^*)$.

Table 2: Computes of (x^*, y^*) for the GBL family.

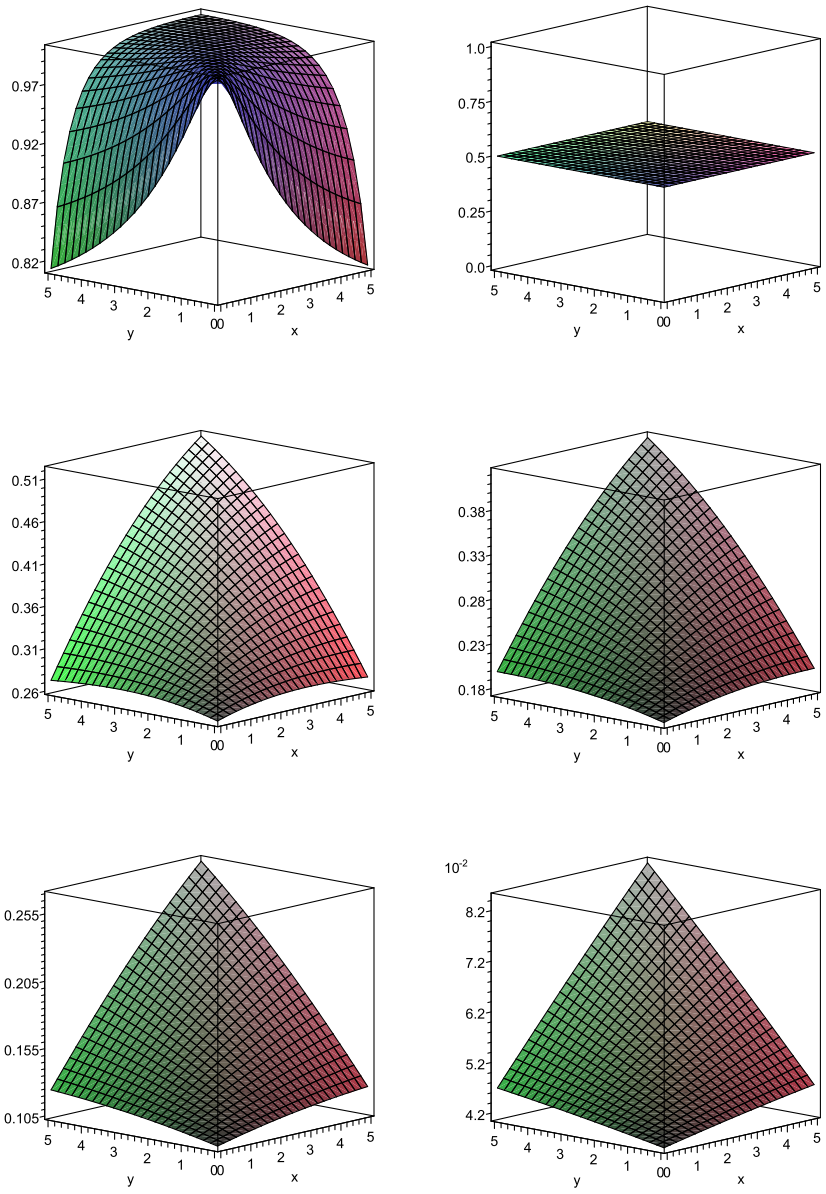

Figure 2: The diagram of $H(X,Y)$ for different values of p. Top left: $p = 1$, Top right: $p = 2$, Middle left: $p = 4$, Middle right: $p = 6$, Bottom left: $p = 10$ and Bottom right $p = 25$.

p	ρ	x^*	y^*	$\Theta(x^*, y^*)$
1	1.000	∞	∞	2.000
2	0.500	1.000	1.000	1.500
3	0.333	0.500	0.500	1.333
4	0.250	0.333	0.333	1.250
5	0.200	0.250	0.250	1.200
6	0.167	0.200	0.200	1.167
7	0.143	0.167	0.167	1.143
8	0.125	0.143	0.143	1.125
9	0.111	0.125	0.125	1.111
10	0.100	0.111	0.111	1.100

5. Some Information measures

In this section, we derive three information measures for GBL family. Also, we study behavior of these measures via a numerical study:

5.1. *Entropy*

If (X, Y) is a random vector with the joint density function $f(x, y)$, the joint entropy for two random variables X and Y is

$$H_e(X, Y) = -E[\log(f(X, Y))].$$

This measure is maximum, when X and Y are independent and if X and Y are dependent random variables, then H_e is a real number.(see, Joe, 1987, 1989 and 1997)

Proposition 5.1. *Let (X, Y) be a random vector with GBL distribution function, then*

$$H_e(X, Y) = \frac{a_2 - 1}{b_1 b_2 a_2}[\ln a_1 + c_1(p)] + \frac{b_2 - 1}{a_1 a_2 b_2}[\ln b_1 + c_1(p)] - c_2(p) - \ln(A),$$

where, $A = a_1 a_2 b_1 b_2 p(p + 1)$ and $c_i(p); i = 1, 2$ are functions of p, see appendices.

Corollary 5.1. *If $a_1 = a_2 = b_1 = b_2 = 1$ then, the entropy of GBL distribution function is as follows:*

$$H_e(X, Y) = \log(p(p + 1)) + \sum_{i=0}^{\infty} \frac{p + 2}{\Gamma(p)\Gamma(i + 1)}[\Gamma(p + i)(\Psi(1 + i) - \Psi(p + i)) + \Gamma(p + i + 1)(\Psi(1 + i) - \Psi(p + i + 1))], \quad (8)$$

where, $\Psi(x) = \partial \ln(\Gamma(x))/\partial x$.

Table 3 and Figure 3 show behavior of $H_e(X, Y)$ with respect to increasing of p, we observe that H_e is positive for $p < 3.55$, zero for $p \cong 3.55$ and negative for $p < 3.55$.

Table 3: Values of $H_e(X,Y)$ for some values of p

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$H_e(X,Y)$	25.12	14.26	10.38	8.29	6.95	6.00	5.27	4.69	4.21
p	1	2	3	**3.55**	4	5	6	7	8
$H_e(X,Y)$	3.81	1.54	0.43	**0**	−0.30	−0.83	−1.26	−1.61	−1.92
p	9	10	11	15	20	25	30	35	40
$H_e(X,Y)$	−2.18	−3.28	−3.89	−4.36	−4.74	−5.05	−5.33	−5.57	−2.41

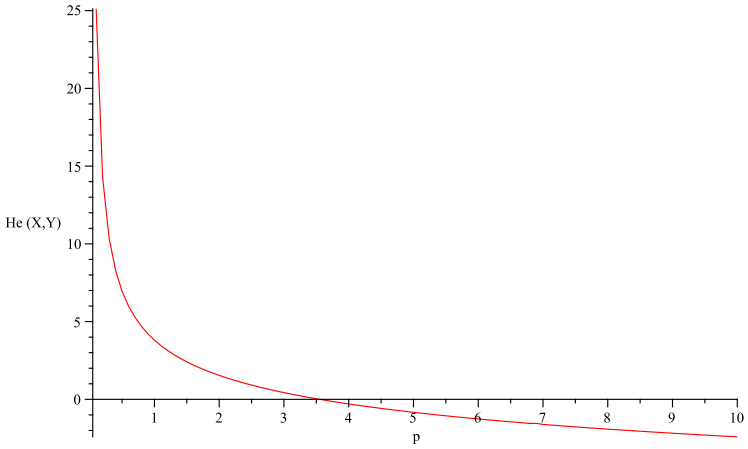

Figure 3: $H_e(X,Y)$ for some values of p

5.2. *Mutual Information*

Mutual information measures are the amount of information that can be obtained about one random variable by observing another. It is important in communication where it can be used to maximize the amount of information shared between sent and received signals. The mutual information of X relative to Y with joint density function $f(x,y)$ and marginal density functions $f_1(x)$ and $f_2(y)$, respectively; given by:

$$I(X,Y) = E\left[\log\left(\frac{f(X,Y)}{f_1(X)f_2(Y)}\right)\right] = H_e(X) + H_e(Y) - H_e(X,Y). \qquad (9)$$

Proposition 5.2. *Let (X,Y) be a random vector with GBL distribution function,*

then

$$I(X,Y) = \frac{a_2-1}{a_2}\left[\ln a_1\left(1-\frac{1}{b_1b_2}\right)+\psi(p)+\gamma-\frac{c_1(p)}{b_1b_2}\right] + \frac{b_2-1}{b_2}\left[\ln b_1\left(1-\frac{1}{a_1a_2}\right)+\psi(p)+\gamma-\frac{c_1(p)}{a_1a_2}\right] + c_3(p)-\log(a_1a_2)-\log(b_1b_2),$$

where $c_1(p)$ and $c_3(p)$ are functions of p, see appendices.

Corollary 5.2. *If $a_1 = a_2 = b_1 = b_2 = 1$ then,*

$$H_e(X) = H_e(Y) = \frac{p+1}{p} - \log(p).$$

Using (8), and (9), we get

$$I(X,Y) = \frac{2p-2}{p} - \log(p^3(p+1)) + \sum_{i=0}^{\infty}\frac{p+2}{\Gamma(p)\Gamma(i+1)}[\Gamma(p+i)(\Psi(1+i)-\Psi(p+i)) + \Gamma(p+i+1)(\Psi(1+i)-\Psi(p+i+1))]. \quad (10)$$

If the component of (X,Y) are independent, then $I(X,Y)$ is zero and conversely when the dependence is maximal, $I(X,Y)$ tends to infinity. Normalizing this index is defined by Joe (1989) as:

$$\delta(X,Y) = \sqrt{1-\exp(-2I(X,Y))}.$$

The measure of δ is confined to the interval $[0,1]$. If X and Y are independent then, $\delta = 0$ and when the dependence is maximal δ achieves to one.
Our numerical results in Table 4 show that when p is decreasing the measures I and δ are increasing and when p is increasing I and δ decreasing. Note that, we observed in sections 3 and 4 that the measures τ, ρ and θ, increase with decreasing of p, mean we have more information and dependence in this case.

Table 4: Values of $I(X,Y)$ and $\delta(X,Y)$ for some values of p

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$I(X,Y)$	1.49	.96	.70	.54	.43	.36	.30	.26	.22
$\delta(X,Y)$	.97	.92	.87	.81	.76	.71	.67	.63	.60
p	1	2	3	4	5	6	7	8	9
$I(X,Y)$	.19	.072	.038	.023	.016	.011	.009	.007	.005
$\delta(X,Y)$	.56	.37	.27	.21	.18	.15	.13	.12	.10
p	10	15	20	25	30	35	40	...	
$I(X,Y)$	.004	.002	.001	.0007	.0005	.0004	.0003	...	
$\delta(X,Y)$	.09	.063	.048	.039	.032	.028	.025	...	

5.3. *Quadratic Mutual Information*

Let X and Y be two random variables with marginal density functions $f_1(x)$ and $f_2(y)$ and joint density function $f(x,y)$, the mutual information between two random variables can be estimated by Kullack-Leibler divergence between the joint density function and the factored marginals. By using quadratic forms of density functions, Xu and Principe (1998), proposed the following distance based on the Cauchy-Schwartz inequality:

$$C(X,Y) = \log \frac{\left(\int_0^\infty \int_0^\infty f^2(x,y)dxdy\right)\left(\int_0^\infty \int_0^\infty f^2(x)f^2(y)dxdy\right)}{\left(\int_0^\infty \int_0^\infty f(x,y)f(x)f(y)dxdy\right)^2}.$$

It is obvious that $C(X,Y) \geq 0$ and $C(X,Y) = 0$ if and only if X and Y are independent. So $C(X,Y)$ is an appropriate measure for the independence of two random variables (minimization of mutual information). Although, it is difficult to prove a strict justification that $C(X,Y)$ is appropriate to measure dependence. We compute this measure for GBL family and then study the behavior of it via a numerical study and drawing it's graph.

Proposition 5.3. *Let (X,Y) be a random vector with GBL distribution function, then,*

$$C(X,Y) = \log(C_1) + \log(C_2) - 2\log(C_3),$$

where,

$$C_1 = Dp^2(p+1)^2 B(2p + \frac{1}{a_2} + \frac{1}{b_2}, 2 - \frac{1}{a_2}, 2 - \frac{1}{b_2}),$$

$$C_2 = Dp^4 B(2p + \frac{1}{a_2}, 2 - \frac{1}{a_2}) B(2p + \frac{1}{b_2}, 2 - \frac{1}{b_2}),$$

$$C_3 = Dp^3(p+1)\left[\frac{B(2 - \frac{1}{a_2}, p + \frac{1}{a_2})\Gamma(1 - p - \frac{1}{a_2})}{\Gamma(p+1)\Gamma(2 - \frac{1}{a_2})} \sum_{i=1}^{\infty} C_4(i) + \frac{B(2p + 1 + \frac{1}{a_2}, -p - \frac{1}{a_2})\Gamma(1 + p + \frac{1}{a_2})}{\Gamma(p+2)\Gamma(2p + \frac{1}{a_2} + 1)} \sum_{i=1}^{\infty} C_5(i)\right],$$

and

$$C_4(i) = \frac{B(2p + \frac{2}{b_2} + i - 1, 2 - \frac{1}{b_2})\Gamma(p + 1 + i)\Gamma(2 - \frac{1}{a_2} + i)}{i!\Gamma(1 - p - \frac{1}{a_2} + i)},$$

$$C_5(i) = \frac{B(p + i + \frac{1}{b_2} - 1, 2 - \frac{1}{b_2})\Gamma(p + 2 + i)\Gamma(2p + 1 + \frac{1}{a_2} + i)}{i!\Gamma(1 + p + \frac{1}{a_2} + i)},$$

$$D = a_2 b_2 a_1^{\frac{1}{a_2}} b_1^{\frac{1}{b_2}}.$$

Corollary 5.3. *If* $a_1 = a_2 = b_1 = b_2 = 1$ *, then,*

$$C(X,Y) = \log\left(\frac{p^2(p+1)^2}{(2p+3)(2p+2)}\right) + 2\log\left(\frac{p^2}{2p+1}\right) - 2\log\left[\frac{p^3\Gamma(-p)}{\Gamma(p+1)}\sum_{i=0}^{\infty}\frac{\Gamma(p+1+i)}{(2p+i)\Gamma(i-p)} + p^3\Gamma(-p-2)\sum_{i=0}^{\infty}\frac{\Gamma(2p+2+i)}{(p+i+1)i!}\right].$$

Under the assumptions of Corollary 5.3, we computed $C(X,Y)$ for some values of p in Table 5, also Figure 4 shows the behavior of $C(X,Y)$ with respect to p. Table 2.4 and Figure 4 show that, increasing p the quadratic mutual information decrease, this agree with behavior of I and δ.

Table 5: Values of $C(X,Y)$ for some values of p

p	1	2	3	4	5	6	7	8	9	10
$C(X,Y)$	.0542	.0227	.0122	.0078	.0044	.0033	.0026	.0020	.0016	.0014

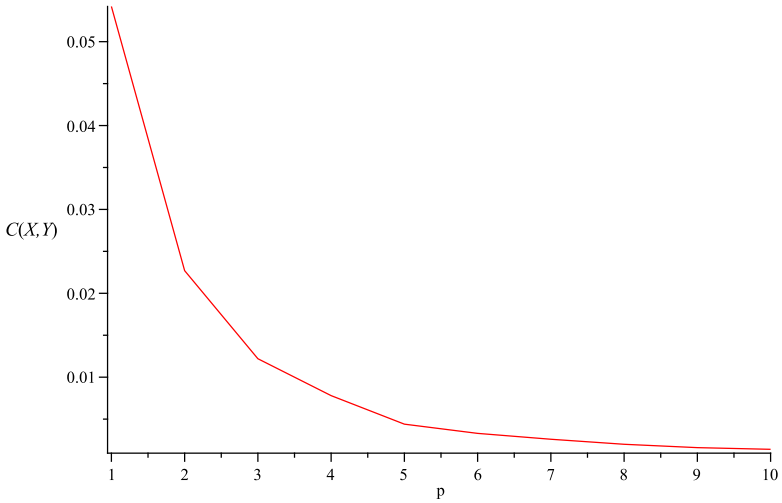

Figure 4: $C(X,Y)$ for some values of p

Remark 5.1. Since the proofs of some propositions are so long, we omit them.

6. Conclusion

In this paper, the dependence structure of GBL family has been studied via dependence coefficients and information coefficients. We show that X and Y in

GBL family have positive dependence and this dependency will be weaker as p goes to be larger. In information measure, term is similar, in fact when p tends to infinity, the measure of information between two random variables tends to be negligible.

References

1. Asadian, N., Amini, M. and Bozorgnia, A. (2008). Aspects of dependence in Lomax distribution. *Commun. Korean Statist. Soc.* **15(2)**, 193-204.
2. Bairamov, I. and Kotz, S. (2002). Dependence structure and symmetry of Hung-Kotz FGM distribution and their extentions. *Metrika.* **56**, 55-72.
3. Bairamov, I. Kotz, S. and Kozubowski, T.J. (2003). A new measure of linear local dependence. *J. Theor. App. Statist.* **37(3)**, 243-258.
4. Barlow, R. E. and Proschan, F. (1981). Statistical Theory of Reliability and Life Testing: Probability Methods. *Holt, Rinehart and Winston, New York.*
5. Blomqvist, N. (1950). On a measure of independence between two random variables. *Ann. Math.Statist. 21:593-600.*
6. Bolbolian, M., Eghbal, N., Amini, M. Azarnoush, H. and Bozorgnia, A. (2009). Aspects of dependence in Cuadras- Auge family. *Commun. Statist. Theory and Methods, To appear.*
7. Bell, C. B. (1962). Mutual information and maximal correlation as measures of dependence. *Ann. Math. Statist.* **33**, 587-597.
8. Caillault, C. and Guegan, D. (2005). Empirical estimation of tail dependence using copulas: application to Asian markets. *Quantitative Finance.* **5**, 489-501.
9. Charpentier, A. and Segers, J. (2006) Lower tail dependence for Archimedian copulas: Characterization and pitfalls. *Insurance: Math. Econ.* **40**, 525-532.
10. Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease insidence. *Biometrika.* **65**, 141-151.
11. Cuadras, C.M. (2006). The importance of being the upper bound in the bivariate family. *SORT.* **30(1)**, 55-84.
12. Cuadras, C. M. and Auge, J. (1981). A continuous general multivariate distribution and its properties. *Commun. Statist. Theory Methods.* **10**, 339-353.
13. Dobric, J. and Schmid, F. (2005) Nonparametric estimation of the lower tail dependence λ_l in bivariate copulas. *J. App. Statist.* **32**, 387-407.
14. Frahm, G. (2006). On the extermal dependence coefficient of multivariate distributions . *Statist.Probab. Letters.* **76**, 1470-1481.
15. Frahm, G. Junker, M., and Schmidt, R. (2005) Estimating the tail- dependence coefficient: Properties and pitfalls. *Math. Econ.* **37** , 80-100.
16. Fredricks, G. A. and Nelsen, R. B. (2007). On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *J. Statist. Plan. Infer.* **137** , 2143-2150.
17. Genest, C.(1987). Frank's family of bivariate distributions.*Biometrika,* 74,3, 549-555.
18. Gupta, R. C. (2003). On some association measure in bivariate distribution functions. *Ann. Math. Statist.* **41**, 713-717.
19. Holland, P.W. and Wang, Y.J.(1987). Dependence function for continiuous bivariate densities. *Commun. Statist. Theory Methods.* **16**, 863-876.

20. Hutchinson, TP. and Lai, CD.(1990). Continuous bivariate distributions, *Emphasising applications. Rumsby Scintific publishing, Adelaide.*
21. Joe, H. (1987) Majorization, randomness and dependence for multivariate distributions. *Ann.Prob.* **15,3**, 1225-1227.
22. Joe, H. (1989) Relative entropy measures of multivariate dependence. *J. Amer. Statist. Assoc.* **84**, 157-164.
23. Joe, H. (1997) Multivariate model and dependence concepts. *Chapman and Hall.*
24. Juri, A. and Wüthrich, M. (2003) Tail dependence from a distributional point of view. *Extremes.* **6**, 213-246.
25. Karlin, S. (1968). Total Positivity. *Stanford University Press. California.*
26. Kotz,S. and Nadarajah, S. (2003). Local dependence function for the elliptically symmetric distributions. *Sankhya*, A, Vol. 65, No.1, pp. 207-223
27. Mari, D. D., and Kotz, S. (2001) Corrolation and dependence.*Imperial College Press.*
28. Nadarajah, S. (2005). Sums, products, and ratios for the bivariate Lomax distribution. *Comput. Statist. and Data Analysis.* **49**, 109-129.
29. Nadarajah, S. Mitov, K. and Kotz, S. (2003). Local dependence functions fo extreme value distributions. *J. App. Statist.* **30(10)**, 1081- 1100.
30. Nayak, T.K. (1987). Multivariate Lomax distribution:Properties and usefulness in reliability theory.*J. App. Probab.*, 24,170-177.
31. Nelsen, R.B. (2006). An introduction to copula. Springer.
32. Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* **84**, 487-493.
33. Ruiz-rivas, C. and Cuadras, C.M. (1988). Inference properties of a one-parameter curved exponential family of distributions with given marginals. *J. Multivariate Anal.* **27(2)**, 447-456.
34. Schweizer, B. and Wolff, E.F. (1981). On Nonparametric Measures of Dependence for Random Variables. *Ann. Statist.* **9**, 879-885
35. Resnik,S.(2006) Heavy-tailed phenomena, probabilistics and statistical modeling.*Springer.*
36. Sankaran,P.G. and Gupta,R.P.(2004). Charaterizations using local dependence function. *Commun. Statist. Theory Methods.* **Vol.33, No.12**, 2959-2974.
37. Shaked, M. (1977). A family of concept of positive dependence for bivariate distributions. *J.Amer. Statist. Assoc.* **72**, 642-650.
38. Sklar, A. (1959). Functions de repartition on n-daimensions et leurs marges. *Publications de lInstitut de Statistique de lUniversite de Paris.* **8**, 229-231.
39. Xu, D. and Principe, J. (1998) Learning From Examples with Quadratic Mutual Information.*Proceedings of 1998 workshop on Neural Ntworks for Signal Processing VIII.*

Vahid Ranjbar.Y
Department of Statistics,
Ordered and Spatial Data Center of Excellence
Faculty of Mathematical Science,
Ferdowsi University of Mashhad,
Mashhad 91755-1159 , Iran.
e-mail : vahidranjbar_y@yahoo.com and va_ra68@stu-mail.um.ac.ir

Mohammad Amini
Department of Statistics,
Ordered and Spatial Data Center of Excellence
Faculty of Mathematical Science,
Ferdowsi University of Mashhad,
Mashhad 91755-1159 , Iran.
e-mail : mamini48@yahoo.com and m-amini@ferdowsi.um.ac.ir

Abolghasem Bozorgnia
Department of Statistics,
Ordered and Spatial Data Center of Excellence
Faculty of Mathematical Science,
Ferdowsi University of Mashhad,
Mashhad 91755-1159 , Iran.

Gholam Reza Mohtashami Borzadaran
Department of Statistics,
Ordered and Spatial Data Center of Excellence
Faculty of Mathematical Science,
Ferdowsi University of Mashhad,
Mashhad 91755-1159 , Iran.

On The Nonparametric Inference In The Exponentiated Models Based On Complete Samples And Extreme Order Statistics

M. Razmkhah and B. Khatib

Department of Statistics, Ferdowsi University of Mashhad

Consider k independent random samples with different sample sizes such that the ith sample comes from the cumulative distribution function (cdf) $F_i = G^{\alpha_i}$, where α_i is a known positive constant and G is an absolutely continuous cdf. This paper shows how one can construct the nonparametric confidence intervals for the quantiles of G. Toward this end, we consider two cases. First we assume that the complete samples are available and then suppose that only the maxima and minima have been observed. In each case exact expressions for the confidence coefficients are derived. Two mentioned procedures are compared via numerical computations and simulation study.

Keywords: Multisampling plan, Order statistics, Proportional reversed hazard model.

1. Introduction and some preliminaries

There are many experiments which have been done in different series such that they are not identical in distribution, but there exists a hierarchical relationship between them. Some of them are corresponding to the increasing sample sizes, such as Olympic games. In some other experiments, the distribution of the interested population shifts by a shock, for example retooling in a factory or management changing in a society. So it is worthwhile to use a multisampling plan to do inference about various characteristics of the baseline distribution such as mean, standard deviation, quantiles and so on. We present our plan as follow:

P: *Consider the data set* $\{X_{i,j}, 1 \leq i \leq k, 1 \leq j \leq n_i\}$, *where* $X_{i,j}$*'s are independent. Moreover for a fixed* i, $X_{i,j}$*'s,* $(1 \leq j \leq n_i)$ *are identically distributed random variables with cdf*

$$F_i(x) = [G(x)]^{\alpha_i}, \tag{1}$$

where $G(\cdot)$*, the first (baseline) cdf, is absolutely continuous and* α_i *is a known positive constant and represent the alteration rate of the* i*th cdf with respect to the baseline cdf, such that* $\alpha_1 = 1$ *and so* $F_1 = G$.

The identity (1) is well-known in the literature as the exponentiated model. This model contains a lot of lifetime distributions such as the power function, exponentiated Weibull, exponentiated Pareto, exponentiated three parameter Burr type XII and other distributions such as the exponentiated normal and exponentiated log-logistic distributions. The exponentiated model has been studied by

several authors, see for example Abdel-Hamid and AL-Hussaini (2009) and Singh *et al.* (2005). The identity (1) is also well-known as the proportional reversed hazard model, see for example Lawless (2003).

For any univariate distribution function G, the population quantile ξ_p of order p, $(0 < p < 1)$ is defined

$$\xi_p = \inf\{x : G(x) \geq p\}.$$

The quantiles are used to construct the concentrations and deviations criteria. The order statistics play an important role in the inferences related to the quantiles; interested reader may refer to the books of Serfling (1980), Arnold *et al.* (1992), David and Nagaraja (2003). In recent years, several articles had been published on non-parametric confidence intervals for quantiles based on usual order statistics. See for example, Krewski (1976), Sathe and Lingras (1981), Hutson (1999) and Zielinski (2005).

Extreme order statistics play an important role in many experiments, particularly in those that are done in different series. For example, the lowest and highest temperature during a week or a month. In this paper, we consider two distinct data set under assumptions **P** and compare them:
(i) we assume that the complete samples are available.
(ii) we suppose that the available data are related to the sample maxima and minima.

The second scheme has been introduced by Razmkhah *et al.* (2008) in which the authors studied the problem of distribution-free confidence intervals based on extreme order statistics in a proportional hazard model, see also Ahmadi and Razmkhah (2007).

The main goal of this paper is to construct confidence intervals for the quantiles of the baseline cdf based on two mentioned data sets. In Sections 2, we use the complete samples to construct nonparametric confidence intervals. Section 3, focuses on extreme order statistics. To compare the results of Sections 2 and 3, the numerical computations and simulation study are given in Section 4.

2. Confidence intervals based on complete samples

Suppose that the assumptions **P** in Section 1 hold and the data set $\{X_{i,j}, 1 \leq i \leq k, 1 \leq j \leq n_i\}$ is available. To construct confidence intervals for ξ_p, the pth quantile of the baseline population G, we arrange $X_{i,j}$'s in ascending order denoting by $Y_{1:N}, \cdots, Y_{N:N}$, where $N = \sum_{i=1}^{k} n_i$. To obtain the main result of this section we need the distributional properties of order statistics in the case of nonidentical component distributions, so we recall the following lemma.

Lemma 2.1. *(David and Nagaraja, 2003) Let X_i, $i = 1, \ldots, k$ be independent and non-identical random variables such that X_i has cdf F_i and denote the ith order statistic of X_j's $(1 \leq j \leq k)$ by $X_{i:k}$. Then*

$$P(X_{i:k} \leq x) = \sum_{r=i}^{k} \sum_{\Gamma_{r,k}} \prod_{s=1}^{r} F_{X_{t_s}}(x) \prod_{s=r+1}^{k} \bar{F}_{X_{t_s}}(x),$$

where the summation index $\Gamma_{r,k}$ *extends over all permutations* $(t_1, \ldots, t_k)$ *of* $\{1, \ldots, k\}$ *for which* $t_1 < \ldots < t_r$ *and* $t_{r+1} < \cdots < t_k$.

Extending Lemma 2.1, we get the following lemma.

Lemma 2.2. *Let* $X_{i,j}$ $(1 \leq i \leq k, 1 \leq j \leq n_i)$ *be independent random variables such that* $X_{i,j}$*'s,* $1 \leq j \leq n_i$*, have cdf* F_i*,* $1 \leq i \leq k$ *and denote the* i*th order statistic of* $X_{i,j}$*'s by* $Y_{i:N}$ *for which* $N = \sum_{i=1}^{k} n_i$*. Then*

$$P(Y_{i:N} \leq x) = \sum_{r=i}^{N} \psi(r, x),$$

where

$$\psi(r, x) = \sum_{\Delta_{r,k}} \prod_{s=1}^{k} \binom{n_s}{j_s} [F_s(x)]^{j_s} [\bar{F}_s(x)]^{n_s - j_s}, \tag{2}$$

for which the summation index $\Delta_{r,k}$ *extends over all integers* $(j_1, \ldots, j_k)$ *that are the nonnegative solutions of equation* $\sum_{s=1}^{k} j_s = r$.

Using Lemma 2.2, we get the following results:

(i) $Y_{i:N}$ is a lower confidence limit for ξ_p, whose confidence coefficient is free of the baseline population and is given by

$$\xi_k(i; p) = P(Y_{i:N} \leq \xi_p) = \sum_{r=i}^{N} \sum_{\Delta_{r,k}} \prod_{s=1}^{k} \binom{n_s}{j_s} p^{\alpha_s j_s} (1 - p^{\alpha_s})^{n_s - j_s}. \tag{3}$$

(ii) $Y_{i:N}$ can be considered as an upper confidence limit for ξ_p, with confidence coefficient $1 - \xi_k(i; p)$.

(iii) $[Y_{i:N}, Y_{j:N}]$, $i < j$, is a two-sided confidence interval for ξ_p, with confidence coefficient

$$\delta_k(i, j; p) = P(Y_{i:N} < \xi_p < Y_{j:N}) = \xi_k(i; p) - \xi_k(j; p), \tag{4}$$

where $\xi_k(i; p)$ is defined in (3).

If p, α_r, n_r $(r = 1, \ldots, k)$ and the desired confidence level γ_0 are specified, we can choose i and j so that $\delta_k(i, j; p)$ achieve to γ_0. Notice that the choice of i and j is not unique, the one that minimizes the expected width of the confidence interval appears reasonable. By Lemma 2.2, we have

$$E(Y_{j:N} - Y_{i:N}) = \sum_{r=i}^{j-1} \int_0^{\infty} [\psi(r, x) + \psi(r, -x)] dx,$$

where $\psi(r, x)$ is defined in (2). Because of the fact that the above expression is a step function of $j - i$, one can choose i and j as close together as possible to achieve a pre-specified confidence level. For given p, α_r's and n_r's, the value of $\delta_k(i, j; p)$ can be computed by numerical methods.

Remark 2.1. For the special case $\alpha_i = 1$, $i \geq 1$, the following results deduce:

(i) $[Y_{i:N}, Y_{j:N}]$ is a confidence interval for ξ_p with confidence coefficient γ_0 if and only if $[Y_{N-j+1:N}, Y_{N-i+1:N}]$ is a confidence interval for ξ_{1-p} with confidence coefficient γ_0, i. e.,

$$P(Y_{i:N} \leq \xi_p \leq Y_{j:N}) = \gamma_0 \Leftrightarrow P(Y_{N-j+1:N} \leq \xi_{1-p} \leq Y_{N-i+1:N}) = \gamma_0.$$

(ii) $Y_{i:N}$ is a lower confidence limit for ξ_p with confidence coefficient γ_0 if and only if $Y_{N-i+1:N}$ is an upper confidence limit for ξ_{1-p} with prediction coefficient γ_0, i. e.,

$$P(Y_{i:N} \leq \xi_p) = \gamma_0 \Leftrightarrow P(Y_{N-i+1:N} \geq \xi_{1-p}) = \gamma_0.$$

3. Confidence intervals based on extreme order statistics

Let $M_i, i = 1, 2, \ldots, k$ be the maximum of a random sample of size n_i; that is,

$$M_i = \max\{X_{i,1}, X_{i,2}, \ldots, X_{i,n_i}\}$$

and $M_i^{'}$ be the corresponding minimum. Then $(M_i^{'}, M_i)$, $1 \leq i \leq k$, are independent random variables with cdfs

$$F_{M_i}(x) = [G(x)]^{\alpha_i n_i}, \quad i = 1, 2, \ldots, k. \tag{5}$$

and

$$F_{M_i^{'}}(x) = 1 - \{1 - [G(x)]^{\alpha_i}\}^{n_i}, \quad i = 1, 2, \ldots, k. \tag{6}$$

Suppose that the available data are only maxima and minima corresponding to the k samples, thus we can construct confidence intervals for ξ_p based on the data set $\mathbf{V} = \{M_1', M_1, \cdots, M_k', M_k\}$. Denote the ith order statistic of the set $\mathbf{V}$ by $V_{i:2k}$, then $[V_{i:2k}, V_{j:2k}]$ can be considered as a confidence interval for ξ_p. To determine the associated confidence coefficient, we first recall the following result from Ahmadi and Razmkhah (2007).

Theorem 3.1. *(Ahmadi and Razmkhah, 2007) Under the assumptions* $\boldsymbol{P}$*, let* $M_r^{'}$ *and* M_r *be corresponding minimum and maximum of the* r*th random sample from distribution* F_r $(r = 1, \ldots, k)$*, respectively, and* $V_{i:2k}$ *be the* i*th order statistic of the set* $\boldsymbol{V}$*. Then*

$$P(V_{i:2k} \leq x) = \sum_{r=i}^{2k} \sum_{m=[\frac{r+1}{2}]}^{min(r,k)} \sum_{A_{r-m,m,k}} \Bigg\{ \prod_{s=1}^{r-m} [F_{t_s}(x)]^{n_{t_s}} \prod_{s=r-m+1}^{j} \big\{1 - [F_{t_s}(x)]^{n_{t_s}} - [1 - F_{t_s}(x)]^{n_{t_s}}\big\} \prod_{s=m+1}^{k} [1 - F_{t_s}(x)]^{n_{t_s}} \Bigg\}, \tag{7}$$

where $[u]$ *stands for the integer part of* u *and* $A_{i_1,i_2,k}$ *extends over all permutations of* $(t_1, \ldots, t_k)$ *from* $\{1, \ldots, k\}$ *such that* $t_1 < \cdots < t_{i_1}$, $t_{i_1+1} < \cdots < t_{i_2}$ *and* $t_{i_2+1} < \cdots < t_k$.

Applying Theorem 3.1, we get the following results:

(i) $V_{i:2k}$ is a lower confidence limit for ξ_p, whose confidence coefficient is free of the baseline population and is given by

$$\begin{aligned}\vartheta_k(i;p) &= P(V_{i:2k} \le \xi_p) \\ &= \sum_{r=i}^{2k} \sum_{m=[\frac{r+1}{2}]}^{min(r,k)} \sum_{A_{r-m,m,k}} \Big\{ p^{\sum_{s=1}^{r-m} \alpha_{t_s} n_{t_s}} \\ &\qquad \prod_{s=r-m+1}^{m} [1 - p^{\alpha_{t_s} n_{t_s}} - (1-p^{\alpha_{t_s}})^{n_{t_s}}] \prod_{s=m+1}^{k} (1-p^{\alpha_{t_s}})^{n_{t_s}} \Big\}. \quad (8)\end{aligned}$$

(ii) $V_{i:2k}$ can be considered as an upper confidence limit for ξ_p, with confidence coefficient $1 - \vartheta_k(i;p)$.

(iii) $[V_{i:2k}, V_{j:2k}]$, $i < j$, is a two-sided confidence interval for ξ_p, with confidence coefficient

$$\gamma_k(i,j;p) = P(V_{i:2k} < \xi_p < V_{j:2k}) = \vartheta_k(i;p) - \vartheta_k(j;p), \quad (9)$$

where $\vartheta_k(i;p)$ is defined in (8).

Remark 3.1. For the special case $\alpha_i = 1$, $i \ge 1$, the following results deduce:

(i) $[V_{i:2k}, V_{j:2k}]$ is a confidence interval for ξ_p with confidence coefficient γ_0 if and only if $[V_{2k-j+1:2k}, V_{2k-i+1:2k}]$ is a confidence interval for ξ_{1-p} with confidence coefficient γ_0, i. e.,

$$P(V_{i:2k} \le \xi_p \le V_{j:2k}) = \gamma_0 \Leftrightarrow P(V_{2k-j+1:2k} \le \xi_{1-p} \le V_{2k-i+1:2k}) = \gamma_0.$$

(ii) $V_{i:2k}$ is a lower confidence limit for ξ_p with confidence coefficient γ_0 if and only if $V_{2k-i+1:2k}$ is an upper confidence limit for ξ_{1-p} with confidence coefficient γ_0, i. e.,

$$P(V_{i:2k} \le \xi_p) = \gamma_0 \Leftrightarrow P(V_{2k-i+1:2k} \ge \xi_{1-p}) = \gamma_0.$$

4. Numerical computations and simulation study

To illustrate the results of this paper, we assume that $k = 3$, $(\alpha_1, \alpha_2, \alpha_3) = (1, 0.8, 1.3)$ and $(n_1, n_2, n_3) = (4, 5, 3)$. The values of $\delta_k(i,j;p)$ and $\gamma_k(i,j;p)$ are computed using Eqs. (4) and (9), respectively. The results are tabulated in Tables 1 and 2, respectively, for selected i, j and p.

As an application the averaged width of the confidence intervals and their variances in exponentiated exponential distribution have been interested and their values are calculated by simulation. Toward this end, the following algorithm has been used:

(i) A random sample of size n_i $(i = 1, 2, 3)$ from $\bar{F}_i(x) = \left(1 - e^{-x}\right)^{\alpha_i}$ (see Eq. (1)) is generated.

(ii) Complete samples are arranged in ascending order and the quantities $L_{i,j} = Y_{j:12} - Y_{i:12}$ are obtained for selected i and j.

(iii) Minimum and maximum of each sample are extracted.

(iv) The minima and maxima are jointly sorted and the quantities $L^*_{i,j} = V_{j:6} - V_{i:6}$ are computed for selected i and j.

(v) The steps **(i)**–**(iv)** are repeated 10^5 times and averaged the values of $L_{i,j}$ and $L^*_{i,j}$, denoting by $\bar{L}_{i,j}$ and $\bar{L^*}_{i,j}$, respectively.

(vii) The variances of $L_{i,j}$ and $L^*_{i,j}$ are calculated by averaging $(L_{i,j} - \bar{L}_{i,j})^2$ and $(L^*_{i,j} - \bar{L^*}_{i,j})^2$, respectively.

The results are presented in Tables 1 and 2 and help us to choose the appropriate confidence intervals for ξ_p.

Note: In Tables 1 and 2, the confidence intervals with the shortest width and confidence coefficient more than 0.95 are indicated by *.

Table 1. Values of $\delta_3(i,j;p)$ for selected i, j and p.

i	j	p=0.2	p=0.3	p=0.4	p=0.5	p=0.6	p=0.7	p=0.8	$L_{i,j}$	$Var(L_{i,j})$
1	12	0.945	0.989	0.998	0.999	0.998	0.986	0.930	3.025	1.552
	11	0.945	0.989	0.998	0.997	0.979	0.912	0.720	2.023	0.559
	10	0.945	0.989	0.995	0.979	0.912	0.740	0.434	1.519	0.308
	9	0.945	0.987	0.981	0.921	0.765	0.496	0.199	1.184	0.196
	8	0.944	0.978*	0.935	0.793	0.547	0.265	0.069	0.934	0.133
	7	0.940	0.945	0.826	0.593	0.319	0.111	0.018	0.733	0.094
2	12	0.762	0.929	0.984	0.997	0.998	0.986	0.930	2.938	1.546
	11	0.762	0.929	0.984	0.994	0.979	0.912	0.720	1.936	0.552
	10	0.762	0.929	0.981	0.977	0.912	0.740	0.434	1.432	0.301
	9	0.762	0.927	0.967*	0.919	0.764	0.496	0.199	1.097	0.189
	8	0.762	0.918	0.921	0.791	0.546	0.265	0.069	0.847	0.126
3	12	0.487	0.777	0.928	0.984	0.995	0.986	0.930	2.839	1.535
	11	0.487	0.777	0.928	0.980	0.977	0.912	0.720	1.837	0.542
	10	0.487	0.777	0.925	0.963*	0.910	0.739	0.434	1.333	0.291
	9	0.487	0.775	0.911	0.905	0.762	0.496	0.199	0.998	0.179
4	12	0.238	0.543	0.797	0.936	0.985	0.984	0.930	2.728	1.524
	11	0.238	0.543	0.797	0.932	0.966*	0.911	0.720	1.726	0.530
	10	0.238	0.543	0.794	0.915	0.899	0.738	0.434	1.223	0.279
5	12	0.088	0.305	0.590	0.823	0.946	0.977	0.929	2.603	1.509
	11	0.088	0.305	0.590	0.819	0.928	0.904	0.719	1.601	0.515
6	12	0.024	0.134	0.359	0.634	0.851	0.951*	0.926	2.459	1.488
	11	0.024	0.134	0.359	0.631	0.833	0.877	0.716	1.458	0.495
7	12	0.005	0.045	0.173	0.407	0.679	0.875	0.912	2.292	1.460

Table 2. Values of $\gamma_3(i,j;p)$ for selected i, j and p.

i	j	p=0.2	p=0.3	p=0.4	p=0.5	p=0.6	p=0.7	p=0.8	$L^*_{i,j}$	$Var(L^*_{i,j})$
1	6	0.945	0.989	0.998	0.999	0.998	0.986	0.930	3.025	1.552
	5	0.945	0.989	0.997*	0.989	0.955	0.855	0.633	1.874	0.553
	4	0.944	0.976*	0.945	0.853	0.687	0.456	0.211	1.156	0.285
	3	0.788	0.630	0.421	0.236	0.106	0.034	0.006	0.441	0.093
2	6	0.619	0.838	0.944	0.984	0.995	0.985	0.930	2.883	1.547
	5	0.619	0.838	0.942	0.974*	0.952*	0.855	0.633	1.732	0.548
	4	0.619	0.825	0.890	0.837	0.683	0.455	0.211	1.014	0.279
3	6	0.157	0.360	0.578	0.763	0.891	0.951*	0.924	2.584	1.544
	5	0.157	0.360	0.576	0.753	0.849	0.821	0.627	1.433	0.548
4	6	0.002	0.013	0.053	0.147	0.311	0.530	0.719	1.869	1.429

From Tables 1 and 2, we observe that:

(i) Confidence intervals based on minima and maxima in two schemes (complete samples and extreme order statistics) provide same results.

(ii) For given p and a fixed confidence level γ_0, one can find a confidence interval with shorter width for ξ_p based on complete samples. For example, a confidence interval with confidence coefficient $\geq 95\%$ for $\xi_{0.5}$, the median of the baseline cdf G, by considering the shortest width criterion, based on extreme order statistics is $(V_{2:6}, V_{5:6})$ with the average width 1.732 and based on complete samples it is $(Y_{3:12}, Y_{10:12})$ with the average width 1.333.

References

Abdel-Hamid, A. H. and AL-Hussaini, E. K. (2009). Estimation in step-stress accelerated life tests for the exponentiated exponential distribution with type-I censoring. *Computational Statistics and Data Analysis*. **53**, 1328–1338.

Ahmadi, J. and Razmkhah, M. (2007). Outer and Inner Confidence Intervals Based on Extreme Order Statistics in a Proportional Hazard Model. *JIRSS*. **6**, 1–16.

Arnold, B. C., Balakrishnan, N., Nagaraja, H. N. (1992). *A First Course in Order Statistics.* Wiley, New York.

David, H. A. and Nagaraja, H. N., (2003). *Order Statistics*, Third edition, John Wiley & Sons, New York.

Hutson, A. D. (1999). Calculating nonparametric confidence intervals for quatiles using fractional order statistics. *J. Appl. Statist.* **26**, 343–353.

Krewski, D., (1976). Distribution-free confidence intervals for quantile intervals. *J. Amer. Statist. Assoc.*, **71**, 420–422.

Lawless, J. L., (2003). *Statistical Models and Methods for Lifetime Data, Second Edition*, John Wiley, New York.

Razmkhah, M., Ahmadi, J. and Khatib, B., (2008). Nonparametric confidence intervals and tolerance limits based on minima and maxima. *Comm. Statist. Theory Methods*, **37**, Issue 10, 1523–1542.

Sathe, Y. S., Lingras, S. R. (1981). Bounds for the confidence coefficient of outer and inner confidence intervals for quantile interval. *J. Amer. Statist. Assoc.*, **76**, 473–475.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Singh, U., Gupta, P. K. and Upadhyay, S.K. (2005). Estimation of parameters for exponentiated-Weibull family under type-II censoring scheme. *Computational Statistics and Data Analysis*. **48**, 509 - 523.

Zielinski, R., Zielinski, W. (2005). Best exact nonparametric confidence intervals for quantiles. *Statistics*, **39**, 67–71.

Semi-parametric Bayesian Models for Transition Longitudinal Data

R. Rikhtehgaran and I. Kazemi

Department of Statistics, University of Isfahan

A common assumption in fitting transition longitudinal data models is normality of stochastic residuals and individual effects. This can be extremely restrictive, making vague most potential features of true distributions. The objective of this paper is to propose a modelling strategy, from a semi-parametric Bayesian perspective, to specify a flexible prior on the space of all possible distribution functions. This is done here by incorporating a Dirichlet process mixture of normal priors as a probability model for the random effects. We also address the role of initial conditions in transition processes, emphasizing on joint modelling of start-up and follow-up responses. We adopt Gibbs sampling techniques to approximate posterior estimates. These important topics are illustrated by testing several hypothetical models in empirical contexts drawn from economic growth studies. We use standard information criteria to select the best fitting model.

Keywords: Dirichlet processes, Dirichlet process mixture, Gibbs sampling, initial conditions, random effects.

1. Introduction

Recently transition longitudinal data models have become increasingly popular in a wide variety of applied researches. These models are used for the analysis of observations that have a serial processes structure which allows for unobserved heterogeneity and lagged responses to examine state dependence (Crouchley and Davies, 2001). It is commonly assumed that heterogeneity effects are uncorrelated with the start-up response (Diggle et al., 2002). Relaxing this restriction in fitting models yields unrealistic results. This is so called the initial conditions problem. A simple solution to this is provided by Wooldridge (2005) who proposes modelling the distribution of the heterogeneity effects conditioned on the first observations. Although this solution is useful to quantify transitions, it will not work in many cases. A special case is when the stochastic process has not been running sufficiently long prior to the sampling period in order to use the stationary distribution as the probability of the first observation, Y_{i0}. To account for this problem, we propose an attractive approach by specifying a reduced form equation for the Y_{i0} and model it jointly with the follow-up observations. The proposed approach effectively combines both likelihoods and handles the initial conditions.

Most researches make simply normality assumption for both stochastic residual terms and individual random effects. This can be extremely restrictive, making vague important features of true distributions. We relax this assumption by adopting a semi-parametric Bayesian method (Escobar, 1994; Muller and Quintana, 2004) by incorporating a Dirichlet process mixture (DPM) of several adaptable priors as probabilities for random effects. The DPM model consists of mixing a Dirichlet process (DP) (Ferguson, 1973), which specifies a prior distribution on the space of all possible distribution functions, with an introduction of a kernel. Such a model assumes that the prior distribution, G, itself is uncertain, but has been drawn from the Dirichlet process. The parameters of DP are a base prior, G_0, which approximates the true non-parametric shape of G, and a precision parameter, M, reflecting the prior belief about how similar G is to G_0.

Despite many attractive features of the DPM model, the practical applications, due to computational difficulties, are limited or even impossible in the frequentist approach. In recent studies, it is shown that how Markov chain Monte Carlo (McMC) methods, such as Gibbs sampling, could be used to overcome these difficulties (Escobar and West, 1998; MacEachern and Muller, 2000). In this paper, we employ the Gibbs sampling algorithm in order to derive all full conditional posterior distributions. We fit DPM models with normal as their base prior for individual random effects of transition longitudinal data by considering initial conditions problem. These concepts are illustrated to an example in economic studies, where the initial conditions and the distribution of random effects are of direct interest. We conclude that inferences regarding the model parameters can crucially depend on the assumptions made and recommend using a DP prior distribution, which is more flexible than the normal. We also report AIC's and BIC's to compare the fitted models.

The remainder of this paper is as follows. In section 2, we briefly review the semi-parametric Bayesian approach and introduce the DPM model. Section 3 considers the specification of transition longitudinal data models with illustration on initial conditions. We first assume DPM as the distribution of random effects, and then present an example, including of unconditional and conditional models which are conditioned on initial values, for the analysis of a real data set. We also assume several distributions for both residual and random effects.

2. The Semiparametric Bayesian Approach

Let G_0 be a probability measure on a (measurable) space $(\tau; \mathcal{A})$ and M be a positive real number. A probability distribution $G(\cdot)$ is distributed according to a DP of base distribution $G_0(\cdot)$ and scale factor M, denoted by $DP(MG_0(\cdot))$, if for any partition $A_1, \cdots, A_k$ of τ and any k, satisfies

$$(G(A_1), \cdots, F(A_k)) \sim D(MG_0(A_1), \cdots, MG_0(A_k))$$

where D is a standard Dirichlet distribution. The total mass parameter M controls the deviation of G from G_0 in a stochastic manner.

In DP models, an important property is that the realizations of a DP are discrete, with probability one. This may lead us using of the stick-breaking procedure (Sethuraman, 1994), defined as $G(\cdot) = \sum_{j=1}^{\infty} \pi_j \delta_{\omega_j}(\cdot)$, where $\omega_j \sim G_0(\cdot)$, $\pi_j = \beta_j \prod_{i=1}^{j-1} (1-\beta_i)$ and $\beta_j \sim Beta(1, M)$. For continuous cases, the prior model is adopted for the unknown distribution F as $F(.) = \sum_{j=1}^{\infty} \pi_j f(.|\omega_j)$.

Let $Y_1, \cdots, Y_n$ be a statistically exchangeable sequence distributed according to the probability density function $f(y \mid \xi)$ where $\xi \in \Xi$ and let $\xi \mid G \sim G$, where $G \sim DP(MG_0)$. Then the density estimation problem, expressed by this hierarchical model, is known as DPM. Escobar (1994) shows that DPM may be simplified by the use of Polya urn representation. To do this, let $\xi_1, \cdots, \xi_N$ be N random samples from G, where $G \sim DP(MG_0)$. It follows that, conditioned on the other ξ's, ξ_i $(i = 1, \cdots, N)$ has the following mixture distribution

$$\xi_i \mid \mathbf{y}, \xi_{(-i)} \propto \sum q_j \delta_{\xi_j} + M q_0 g_0(\xi_i) f(y_i|\xi_i) \tag{1}$$

where $\xi_{(-i)} = \left(\xi_1^{'}, ..., \xi_{i-1}^{'}, \xi_{i+1}^{'}, ...\xi_N^{'}\right)^{'}$, $\delta_{\xi_j}$is a degenerate distribution with point mass at ξ_j, $q_j = f(y_i \mid \xi_j)$, $j = 1, ..., i-1, i+1, ...N$, and

$$q_0 = \int f(y_i \mid \xi)\, g_0(\xi) d\xi. \tag{2}$$

We note that several ξ_i's might have the same value such that the number of distinct values of ξ_i, denoted by K, is less than N. The scaling coefficient M tunes the number of clusters K. For large N, it is shown that $E[K \mid M, N] \simeq Mlog\left(1 + \frac{N}{M}\right)$, for more information you can see Antoniak (1974) and Dorazio (2009). As M tends to zero, most of the samples ξ_i share the same value, whereas when M tends to infinity, the ξ_i are almost *i.i.d.* samples from G_0. Thus M plays a critical role in smoothing the distribution.

3. The Specification of Transition Longitudinal Data Models

We now apply the DPM in fitting an longitudinal model with focus on the role of initial conditions. We suggest here to consider a reduced form equation for the initial response and impose a random effect correlated with the follow-up model equation. The estimation method would then effectively combine all information contained in the transition process. That is, for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, we consider

$$Y_{i0} = \mathbf{x}_{i0}'\beta_0 + \varphi\alpha_i + \varepsilon_{i0}$$
$$Y_{it} = \gamma Y_{i,t-1} + \mathbf{x}_{it}'\beta + \alpha_i + \varepsilon_{it}$$

These model equations can be shown in the vector form

$$\mathbf{Y}_i = \widetilde{\mathbf{X}}_i\theta + \mathbf{z}\alpha_i + \varepsilon_i, \quad i = 1, \cdots, N, \tag{3}$$

where $\mathbf{Y}_i = (Y_{i0}\ Y_{i1} \cdots Y_{iT})'$, $\theta = (\beta_0'\ \gamma\ \beta')'$, $\mathbf{z} = (\varphi\ \mathbf{e}_T')'$where parameters β and β_0 are vectors with $p+1$ dimensions including intercept parameters, $\mathbf{e}_T$ is a unit vector of order T, and $\widetilde{\mathbf{X}}_i$ is defined according to θ.

In fitting longitudinal models, the usual assumptions are both the individual effects and the residual terms are normally distributed. This can be an unrealistic assumption in many situations. We now relax this assumption for the individual effects by applying DPM, where the base distribution is assumed normally distributed. Consider the following hierarchical model

$$\begin{aligned} &\mathbf{Y}_i \mid \alpha_i \overset{ind}{\sim} N_{T+1}\left(\widetilde{\mathbf{X}}_i\theta + \mathbf{z}\alpha_i, \sigma_\varepsilon^2\mathbf{I}_{T+1}\right) \\ &\alpha_i \mid G \overset{iid}{\sim} G \\ &G \mid M \sim DP\left(MG_0\right), G_0 \sim \quad N\left(0, \sigma_\alpha^2\right) \end{aligned} \tag{4}$$

The priors are assumed as follows; the inverse gamma distribution with the hyperparameters (τ_1, τ_2) for σ_ε^2 and (δ_1, δ_2) for σ_α^2. The multivariate normal distribution with the mean vector θ_0 and the covariance matrix $\mathbf{\Lambda}$ for θ, and the normal distribution with mean φ_0 and variance σ_φ^2 for φ.

Data analysis are conducted using the Gibbs sampling technique which is a particular McMC algorithm. The Gibbs sampler proceeds by iterative simulation from the full conditional posterior distributions of each unknown stochastic parameter given the current values of all other model parameters and the observations. In below, we apply the Gibbs sampling algorithm, in order to find full conditional posteriors.

To find the conditional posterior $\alpha_i|\theta, \sigma_\alpha^2, \alpha_{(-i)}, \mathbf{Y}$ $(i = 1, \cdots, N)$, let $\alpha_{(-i)}$ be a vector of α's, after removing α_i. It can readily be shown that

$$g_0\left(\alpha_i\right) f\left(\mathbf{y}_i \mid \alpha_i\right) = \phi\left(\alpha_i \mid \mu_i^*, \sigma_*^2\right)\ \phi_{T+1}\left(\mathbf{y}_i \mid \widetilde{\mathbf{X}}_i\theta, \sigma_\varepsilon^2\mathbf{\Omega}^{-1}\right) \tag{5}$$

where, for $\sigma_c^2 = \sigma_\varepsilon^2 + \left(T + \varphi^2\right)\sigma_\alpha^2$ and $\psi = \sigma_\alpha^2/\sigma_c^2$, we have

$$\mu_i^* = \psi\mathbf{z}'\left(\mathbf{y}_i - \widetilde{\mathbf{X}}_i\theta\right),\ \sigma_*^2 = \sigma_\varepsilon^2\psi,\ \mathbf{\Omega} = \mathbf{I} - \psi\mathbf{z}\mathbf{z}'. \tag{6}$$

Integrating out the α_i's, the probability q_0 , given in (2) is simplified as

$$\int g_0\left(\alpha_i\right) f\left(\mathbf{y}_i \mid \alpha_i\right) d\alpha_i = \phi_{T+1}\left(\mathbf{y}_i \mid \widetilde{\mathbf{X}}_i\theta, \sigma_\varepsilon^2\mathbf{\Omega}^{-1}\right) \tag{7}$$

Using the mixture distribution (1), the simulation of α_i's is based on the following scheme. With the probability proportional to $q_j = f\left(\mathbf{Y}_i \mid \alpha_j\right) = \phi_{T+1}\left(\widetilde{\mathbf{X}}_i\theta + \mathbf{z}\alpha_j, \sigma_\varepsilon^2\mathbf{I}_{T+1}\right)$, we draw α_i according to δ_{α_j}, and with the probability

proportional to Mq_0, we draw α_i according to $N\left(\mu_i^*, \sigma_*^2\right)$. After some algebra is done, we can show that

$$\sigma_\varepsilon^2|\theta, \sigma_\alpha^2, \varphi, \alpha, \mathbf{y} \sim IG\left(s_1, s_2\right), \tag{8}$$

where $s_1 = \tau_1 + N(T+1)/2$, $s_2 = \tau_2 + \frac{1}{2}\sum_{i=1}^{N}\left(r_i - \mathbf{z}\alpha_i\right)'\left(\mathbf{r}_i - \mathbf{z}\alpha_i\right)$, and $r_i = \mathbf{y}_i - \widetilde{\mathbf{X}}_i\theta$ are fitted residuals. For the regression coefficients, we find that

$$\theta|\sigma_\varepsilon^2, \sigma_\alpha^2, \varphi, \alpha, \mathbf{Y} \sim N_{2p+3}\left(a_1 b_1, a_1\right) \tag{9}$$

where $a_1 = \left(\Lambda^{-1} + NW_{\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}i}\right)^{-1}$, $b_1 = \theta_0'\Lambda^{-1} + \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{N}\left(\mathbf{Y}_i - \mathbf{z}\alpha_i\right)'\widetilde{\mathbf{X}}_i$, and $W_{\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}} = \frac{1}{N\sigma_\varepsilon^2}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i$ is the within-variation matrix. We also obtain

$$\varphi|\sigma_\varepsilon^2, \sigma_\alpha^2, \theta, \alpha, \mathbf{Y} \sim N\left(a_2 b_2, a_2\right) \tag{10}$$

where $a_2 = \left(\frac{1}{\sigma_\varphi^2} + \frac{1}{\sigma_\varepsilon^2}SSA\right)^{-1}$, $b_2 = \frac{1}{\sigma_\varphi^2}\varphi_0 + \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{N} r_{i0}\alpha_i$, and SSA is the sum of squared of random effects.

By the use of West, Muller, and Escobar (1994) approach, the subjects will be grouped into clusters so that similar α_i's discriminate in the same group. Consequently, there will be K unique individual effects, denoted by η_l, $l = 1, ..., K$. Thus, the full posterior of σ_α^2 conditioned on α_i's, implies that the α_i's are known for us. This also imply K and the η_l's are all known. Consequently, the K independent variables η_l's are distributed as $N\left(0, \sigma_\alpha^2\right)$. Now, it can be readily shown that

$$\sigma_\alpha^2|\theta, \sigma_\varepsilon^2, \varphi, \alpha, \mathbf{Y} \sim IG\left(s_3, s_4\right) \tag{11}$$

where $s_3 = \delta_1 + K/2$, $s_4 = \delta_2 + \frac{1}{2}SSB$, and SSB is the sum of squared of η_l's.

The Gibbs sampler proceeds by simulating a sequence of the above conditional random variables. Following a sufficient burn-in period, the Gibbs sequence converges to a stationary distribution which is the target distribution that we are trying to simulate. The samples then can be used in the computation of any feature of either marginal posterior distribution.

3.1. *An Empirical Study*

The data are regular observations in five years $t = 1965, 1970, 1975, 1980$, and 1985 for 94 countries which are taken from Islam (1995). In this study, the growth convergence equation is derived from assumptions about a production function and inclusion in the specification of the savings rate, s, and the population growth rate, p. Suppose y_{it} to be the logarithm of per capita GDP for country i at time t . We consider the following model

$$y_{i,t} = \lambda + \gamma y_{i,t-1} + \beta x_{it} + \alpha_i + \varepsilon_{it}$$
$$y_{i0} = \lambda_0 + \beta_0 x_{i0} + \varphi\alpha_i + \varepsilon_{i0}$$

where $\mathbf{x}_{it}' = (1, x_{i,t})$, $\beta' = (\lambda, \beta)$, $\mathbf{x}_{i0}' = (1, x_{i,0})$, $\beta_0' = (\lambda_0, \beta_0)$ and $x_{i,t} = log(s_{i,t}) - log(p_{i,t})$. Also, α_i is a country-specific effect and ε_{it} is the residual term. We fit

the conditional transition model, which ignores the initial conditions, and the full unconditional model by making various assumptions on the distributions of residuals and country effects. We then use AIC and BIC to select the best fitted model. The specification of models are represented as:

Model 1: Residuals and random effects are both normally distributed with zero means and variances σ_ε^2 and σ_α^2, respectively.
Model 2: Residuals and random effects both have t distributions, each with zero location, scale parameters σ_ε^2 and σ_α^2, degrees of freedom ν_ϵ and ν_α, respectively.
Model 3: The distribution of residuals is assumed t, while random effects are normally distributed.
Model 4: DPM models are considered for random effets, based on the specification of Section 3.
Model 5. The residual terms have t distributions, while the random effects obey DPM with the base normal distribution.

In fitting models, the hyper-parameters for inverse-gamma priors are assumed both equal to 0.01; for φ the location parameter and the precision parameter set to zero and 0.001, respectively; for regression coefficients, we assume each normally distributed with zero means and 0.001 precisions. For the representation of DP, the stick-breaking is considered. Three values of the parameter M are chosen to reflect departures from normality assumption. In our investigation, we set $M = 2.5$, 20 and 1000, to show small, moderate, and large departures, respectively. The corresponding values for parameter K are, respectively, $K = 4$, 15 and 39.

The Bayes estimates of model parameters are obtained by using WinBugs software. After 5000 burn-in, the Gibbs sampler run for 10000 iterations to ensure that the convergence is achieved. The results are shown in Tables 1 and 2.

It is seen that AIC's and BIC's are lower in the unconditional models, emphasizing the better fitted models when acknowledging the role of initial conditions. The estimates of γ are statistically significant and positively much less than one, resulting in strong evidence for state dependence and suggesting that the countries with low initial GDP per capita values are growing faster than those with high values. Among Model 1 through Model 5, we may conclude that the Student's t is suitable for the residuals. Finally, values of AIC and BIC show that Model 5 is the best fitted model.

References

1. Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics* **2**, 1152–1174.
2. Crouchley, R. and Davies R.B. (2001). A Comparison of GEE and Random Effects Models for Distinguishing Heterogeneity, Nonstationarity and State Dependence in a Collection of Short Binary Event Series. *Statistical Modelling: An International Journal*, **1**(**4**), 271-285.

Table 1. Parameter estimations of unconditional models.

	Model1	Model2	Model3	Model4 K=4	Model4 K=15	Model4 K=39	Model5 K=15
λ_0	7.056	7.069	7.056	6.918	7.00	7.090	7.088
	(0.085)	(0.080)	(0.075)	(0.353)	(0.275)	(0.263)	(0.387)
β_0	0.701	0.705	0.719	0.616	0.624	0.567	0.618
	(0.074)	(0.078)	(0.066)	(0.027)	(0.034)	(0.042)	(0.027)
φ	8.446	6.617	6.811	9.380	7.680	9.006	7.171
	(1.185)	(0.856)	(0.950)	(1.879)	(0.965)	(1.463)	(0.776)
λ	1.055	1.262	1.241	0.966	1.137	1.021	1.229
	(0.109)	(0.128)	(0.130)	(0.128)	(0.112)	(0.113)	(0.088)
β	0.191	0.217	0.216	0.176	0.191	0.175	0.203
	(0.013)	(0.015)	(0.016)	(0.015)	(0.013)	(0.014)	(0.012)
γ	0.859	0.829	0.831	0.869	0.846	0.864	0.835
	(0.015)	(0.018)	(0.018)	(0.018)	(0.014)	(0.016)	(0.012)
ν_e	-	4.300	4.610	-	-	-	7.055
		(1.266)	(1.571)				(2.793)
ν_α	-	44.920		-	-	-	-
		(48.520)					
σ_e^2	0.018	0.010	0.010	0.022	0.018	0.022	0.013
	(0.001)	(0.001)	(0.002)	(0.073)	(0.001)	(0.002)	(0.001)
σ_α^2	0.007	0.010	0.010	0.028	0.016	0.014	0.020
	(0.002)	(0.003)	(0.003)	(0.073)	(0.009)	(0.007)	(0.012)
AIC	-659	-829	-825	-542	-626	-551	-775
BIC	-627	-789	-789	-510	-621	-519	-743

Table 2. Parameter estimations of conditional models.

	Model1	Model2	Model3	Model4 K=4	Model4 K=15	Model4 K=39	Model5 K=15
λ	0.588	0.617	0.607	0.510	0.544	0.374	0.697
	(0.087)	(0.095)	(0.105)	(0.098)	(0.090)	(0.148)	(0.078)
β	0.145	0.150	0.149	0.135	0.141	0.128	0.158
	(0.014)	(0.015)	(0.016)	(0.012)	(0.013)	(0.011)	(0.013)
γ	0.924	0.921	0.922	0.938	0.930	0.946	0.907
	(0.012)	(0.012)	(0.015)	(0.010)	(0.012)	(0.009)	(0.011)
ν_e	-	5.549	5.580	-	-	-	3.956
		(1.855)	(1.969)				(1.042)
ν_α	-	29.790		-	-	-	-
		(39.670)					
σ_e^2	0.017	0.010	0.010	0.018	0.017	0.019	0.009
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
σ_α^2	0.004	0.004	0.004	0.030	0.011	0.089	0.015
	(0.001)	(0.001)	(0.001)	(0.067)	(0.008)	(0.207)	(0.009)
AIC	-377	-526	-527	-309	-366	-329	-566
BIC	-341	-460	-487	-273	-330	-293	-524

Note: [a] The numbers in paranteses are Bayesain standard deviations.

3. Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L. (2002). *Analysis of Longitudinal Data*. 2nd, Oxford University Press.
4. Dorazio, R.M. (2009). On selecting a prior for the precision parameter of Dirichlet process mixture models. *Journal of Statistical Planning and Inference* **139**, 3384–3390.
5. Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *The Journal of American Statistical Association*, **89**(**425**), 268-277.
6. Escobar, M.D. and West, M. (1998). Computing Bayesian nonparametric hierarchical models. *A revision of the paper of the authors in (1992).* Discussion Paper 92-A20. Duke Univarsity. ISDS.
7. Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics***1**, 209–230.
8. Islam, N. (1995). Growth Empirics: A Panel Data Approach. *Quarterly Journal of Economics*, **110**, 1128-70.
9. MacEachern, S. N. and Muller, P. (2000). Efficient MCMC schemes for robust model extensions using encompassing Dirichlet process mixture models. In *Robust Bayesian Analysis,* F. Ruggeri and D. Rıos-Insua (eds), 295–316. New York: Springer-Verlag.
10. Muller, P. and Quintana, F. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**, 95–110.
11. Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**, 639–650.
12. West, M., Muller, P., and Escobar, M.D. (1994). Hierarchical priors and mixture models, with applications in regression and density estimation. In *Aspects of Uncertainty: A Tribute to D.V. Lindley(Eds. A.F.M Smith and P.R. Freeman).* New York: Wiley.
13. Wooldridge, J.M. (2005) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* **20**, 39–54.

On Restricted Semiparametric Models

M. Roozbeh[1], M. Arashi[2] and H. A. Niroumand[1]

[1] *Department of Statistics, Ferdowsi University of Mashhad*

[2] *Faculty of Mathematics, Shahrood University of Technology*

In this approach we consider a semiparametric regression model, for multicollinear systems. Under a complicated situation, first of all we assume that the parameter space is restricted, and propose necessary and sufficient conditions for the superiority of the new estimator over the restricted least-squares estimator. Furthermore, a nonparametric estimation after estimation of linear part is added for detecting the efficiency of the difference-based approach.

Keywords: Differencing estimator, Least squares, Linear restrictions, Multicollinearity, Partial linear model, Ridge estimator.

1. Introduction

Consider a nonparametric regression model

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, ..., n \tag{1}$$

where the y_i's are the observations, $f(\cdot)$ is an unknown function and the ϵ_i's are independent and identically distributed random variables with zero mean and variance σ^2 and the t's have bounded support. All we know about $f(\cdot)$ is that it's first derivative is bounded by a constant, say L. Usually one fits the function $f(\cdot)$ first and then estimates the variance σ^2 from residual sum of squares.

The second class of estimators use differences that aim to remove the trend in the data that arises from the function $f(\cdot)$. Such methods do not require an estimator of the function $f(\cdot)$ and are often called difference-based estimators. Provided that $f(\cdot)$ is differentiable and the t ordinates are closely spaced, it is possible to remove the effect of the function $f(\cdot)$ by differencing the data appropriately.

The idea of differencing to remove the nonparametric effect in nonparametric and semiparametric regression models is not new. In the partial linear model it is examined by Ahn and Powell (1993). In a pure nonparametric regression setting the idea of differencing has a longer history have been used to obtain estimators of the residual variance. The difference-based estimation procedure is optimal in the sense that the estimator of the linear component is asymptotically efficient and

the estimator of the nonparametric component is asymptotically minimax rate optimal for the partial linear model (Wang *et al.* 2007).

The gist of this approach is to deal with semiparametric models under differencing methodology, for multicollinearity settings. Multicollinearity is defined as the existence of nearly linear dependency among column vectors of the design matrix $\mathbf{X}$ in the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{b} + \boldsymbol{\epsilon}$. The existence of multicollinearity may lead to wide confidence intervals for individual parameters or linear combination of the parameters and may produce estimates with wrong signs, etc. The best way of explaining the existence and structure of multicollinearity is to look at the eigenvalues and eigenvectors of $\mathbf{X}'\mathbf{X}$. If $\mathbf{X}'\boldsymbol{X}$ is ill-conditioned with a large condition number a ridge regression estimator can be used to estimate bb. In this paper, we will examine a biased estimation technique when the matrix $\mathbf{X}'\mathbf{X}$ appears to be ill-conditioned in the partial linear model.

2. Model

Let $(y_1, x_1, t_1), \cdots, (y_n, x_n, t_n)$ be the observations that follow the partial linear model given by

$$y_i = \boldsymbol{x}_i\boldsymbol{b} + f(t_i) + \epsilon_i, \quad i = 1, ..., n \tag{2}$$

where $f(\cdot)$ is as before, $\boldsymbol{x}_i's$ are known p-vectors for $i = 1, ..., n$ and $\boldsymbol{b} = (\beta_1, ..., \beta_p)'$ is a vector of unknown parameters and we assume that in general, $\boldsymbol{\epsilon}$ is a n-vector of disturbances distributed with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2\boldsymbol{I}_n$. Partial linear models are more flexible than standard linear models since they have a parametric and a nonparametric component. They can be a suitable choice when one suspects that the response y linearly depends on x, but that it is nonlinearly related to t. In model (2), Yatchew (1997) concentrates on estimation of the linear component and used differencing to eliminate bias induced from the presence of the nonparametric component. Wang et al. (2007) used higher order differences for optimal efficiency in estimating the linear part by using a special class of difference sequences. They noted that, although the differences are correlated, the correlation should be ignored and the linear regression coefficient vector should be estimated by the ordinary least squares estimator instead of a generalized least squares estimator which takes into account the correlations among the differences. If the correlation structure is incorporated in the estimation, the resulting generalized least squares estimator will not be optimal. To motivate the form of the difference-based ridge estimator first rewrite model (2) in matrix/vector notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{f}(t) + \boldsymbol{\epsilon}, \tag{3}$$

where $\boldsymbol{y} = (y_1, ..., y_n)', \boldsymbol{f}(t) = (f(t_1), ..., f(t_n))', \boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)$ and $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)'$ is the $n \times p$ matrix. Let $\boldsymbol{d} = (d_0, ..., d_m)$ be a $m + 1$ vector, where m is the order of differencing and $d_0, ..., d_m$ are differencing weights satisfying the

conditions

$$\sum_{j=0}^{m} d_j = 0, \quad \sum_{j=0}^{m} d_j^2 = 1. \tag{4}$$

A differencing matrix denoted by $\boldsymbol{D}$ is a $(n-m)\times n$ known matrix with the elements satisfying (4) (see Yatchew, 2003 for some examples). Imposing the differencing matrix to the model (3), permits direct estimation of the parametric effect. In particular, it takes

$$\boldsymbol{Dy} = \boldsymbol{DXb} + \boldsymbol{Df}(t) + \boldsymbol{D\epsilon}. \tag{5}$$

Since the data have been reordered so that the $\boldsymbol{X}$'s are close, the application of the differencing matrix $\boldsymbol{D}$ in model (5) removes the nonparametric effect in large samples (Yatchew 2000). Thus, the underlying model is rewritten as

$$\boldsymbol{y} \doteq \boldsymbol{Xb} + \boldsymbol{e}, \tag{6}$$

where $\boldsymbol{y} = \boldsymbol{Dy}$, $\boldsymbol{X} = \boldsymbol{DX}$, $\boldsymbol{e} = \boldsymbol{D\epsilon}$.

Differencing allows one to perform inferences on $\boldsymbol{b}$ as if there were no nonparametric component f in the model (2). Once $\boldsymbol{b}$ is estimated, a variety of nonparametric techniques could be applied to estimate $f(\cdot)$ as if $\boldsymbol{b}$ were known.

3. Proposed Estimator

It can be totally accepted (Yatchew, 1997) that adopting the linear model (6), the unbiased estimator of $\boldsymbol{b}$ is the following difference-based estimator given by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{D}} = {\boldsymbol{C_D}}^{-1}\boldsymbol{X'y}, \quad \boldsymbol{C_D} = \boldsymbol{X'X}. \tag{7}$$

It is observed from (7) that the properties of the difference-based estimator of $\boldsymbol{b}$ depends heavily on the characteristics of the information matrix $\boldsymbol{C_D}$. If the $\boldsymbol{C_D}$ matrix is ill-conditioned (near dependency among various columns of $\boldsymbol{C_D}$), then the $\hat{\boldsymbol{\beta}}_{\boldsymbol{D}}$ produce unduly large sampling variances. Moreover, some of the regression coefficients may be statistically insignificant with wrong sign and meaningful statistical inference become difficult for the researcher. As a remedy following of Hoerl and Kennard (1970), it can be suggested to use the following estimator namely difference-based ridge estimator

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{D}}(k) = \boldsymbol{T}_k\hat{\boldsymbol{\beta}}_{\boldsymbol{D}}, \quad \boldsymbol{T}_k = (k{\boldsymbol{C_D}}^{-1} + \boldsymbol{I}_p)^{-1}, \tag{8}$$

where $k \geq 0$ is the shrinking parameter.

Now consider the linear non-stochastic constraint

$$\boldsymbol{Rb} = \boldsymbol{r}, \tag{9}$$

for a given $q \times p$ matrix $\boldsymbol{R}$ with rank $q < p$ and a given $q \times 1$ vector $\boldsymbol{r}$. Subject to the linear restriction (9), the restricted difference-based estimator is given by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{RD}} = \hat{\boldsymbol{\beta}}_{\boldsymbol{D}} - {\boldsymbol{C_D}}^{-1}\boldsymbol{R'}(\boldsymbol{R}{\boldsymbol{C_D}}^{-1}\boldsymbol{R'})^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{D}} - \boldsymbol{r}). \tag{10}$$

By Swamy et al. (1978) and Swamy and Mehta (1977) we obtain restricted difference-based ridge estimator to improve the difference-based estimator by minimizing the sum of squared residuals given by

$$\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k) = \boldsymbol{D}(k) - \boldsymbol{C_D}(k)^{-1}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C_D}(k)^{-1}\boldsymbol{R}']^{-1}[\boldsymbol{R}\boldsymbol{D}(k) - \boldsymbol{r}], \tag{11}$$

where $\boldsymbol{C_D}(k) = \boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p$ and $\boldsymbol{D}(k) = \boldsymbol{C_D}(k)^{-1}\boldsymbol{X}'\boldsymbol{y}$ for $k \geq 0$.

Then it is easy to see that the $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$ and $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$ is restricted with respect to $\boldsymbol{R}\boldsymbol{b} = \boldsymbol{r}$. It is also clear that for $k = 0$, we get $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(0) = \hat{\boldsymbol{\beta}}_{\mathbf{RD}}$.

4. Evaluation of Risk Functions

In this section, we calculate the risk function for the proposed estimator given in previous section under quadratic loss function.

For driving the risk function of $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$, according to Roozbeh et al. (2009), $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$ can be written as follows:

$$\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k) = \boldsymbol{N_D}(k)\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{N_D}(k)\boldsymbol{C_D}(k)\boldsymbol{b}_0 + \boldsymbol{b}_0, \tag{12}$$

where, $\boldsymbol{b}_0 = \boldsymbol{R}'(\boldsymbol{R}\boldsymbol{R}')^{-1}\boldsymbol{r}$ and

$$\boldsymbol{N_D}(k) = \boldsymbol{C_D}(k)^{-1} - \boldsymbol{C_D}(k)^{-1}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C_D}(k)^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{C_D}(k)^{-1}. \tag{13}$$

Now, we can calculate the bias and covariance matrix of $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$ by using equation (12) as follows:

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k) - \boldsymbol{b}] &= \boldsymbol{N_D}(k)\boldsymbol{C_D}\boldsymbol{b} - \boldsymbol{N_D}(k)\boldsymbol{C_D}(k)\boldsymbol{b}_0 + \boldsymbol{b}_0 - \boldsymbol{b} \\ &= \boldsymbol{N_D}(k)\boldsymbol{C_D}(k)(\boldsymbol{b} - \boldsymbol{b}_0) - k\boldsymbol{N_D}(k)\boldsymbol{b}_0 + \boldsymbol{b}_0 - \boldsymbol{b} \\ &= -k\boldsymbol{N_D}(k)\boldsymbol{b}, \end{aligned} \tag{14}$$

With direct calculation using $\boldsymbol{C_D} = \boldsymbol{C_D}(k) - k\boldsymbol{I}_p$, we conclude that $\boldsymbol{N_D}(k)\boldsymbol{C_D}\boldsymbol{N_D}(k) = \boldsymbol{N_D}(k) - k\boldsymbol{N_D}(k)^2$ and in particular, $\boldsymbol{N_D}(0)\boldsymbol{C_D}\boldsymbol{N_D}(0) = \boldsymbol{N_D}(0)$ when $k = 0$.

Therefore, the MSE of $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$ is

$$Cov[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)] = \sigma^2\boldsymbol{N_D}(k)\boldsymbol{C_D}\boldsymbol{N_D}(k) = \sigma^2\boldsymbol{N_D}(k) - k\sigma^2\boldsymbol{N_D}(k)^2. \tag{15}$$

Then the bias and covariance matrix of $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$ is obtained by letting $k = 0$ in (14) and (15) as follow:

$$E(\hat{\boldsymbol{\beta}}_{\mathbf{RD}} - \boldsymbol{b}) = \mathbf{0}, \tag{16}$$

$$Cov(\hat{\boldsymbol{\beta}}_{\mathbf{RD}}) = \sigma^2\boldsymbol{N_D}(0). \tag{17}$$

So the risk function of the estimators under study can be expressed by

$$MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k), \boldsymbol{b}] = \sigma^2\boldsymbol{N_D}(k) - k\sigma^2\boldsymbol{N_D}(k)^2 + k^2\boldsymbol{N_D}(k)\boldsymbol{b}\boldsymbol{b}'\boldsymbol{N_D}(k), \tag{18}$$

and

$$MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}, \boldsymbol{b}] = \sigma^2\boldsymbol{N_D}(0). \tag{19}$$

Theorem 4.1. *There exists at least a $k^* > 0$ such that $\hat{\boldsymbol{\beta}}_{\boldsymbol{RD}}(k^*)$ dominates $\hat{\boldsymbol{\beta}}_{\boldsymbol{RD}}$ in the sense of MSE.*

Proof. It is enough to show that there exists $k^* > 0$ such that $MSE(\hat{\boldsymbol{\beta}}_{\mathbf{RD}}, \boldsymbol{b}) - MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k^*), \boldsymbol{b}] > 0$. The partial derivative of (18) with respect to k is

$$\frac{\partial MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k), \boldsymbol{b}]}{\partial k} = -k^2 \boldsymbol{N_D}(k)^2 \boldsymbol{b}\boldsymbol{b}' \boldsymbol{N_D}(k) - k^2 \boldsymbol{N_D}(k) \boldsymbol{b}\boldsymbol{b}' \boldsymbol{N_D}(k)^2,$$

then $\left.\frac{\partial MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k), \boldsymbol{b}]}{\partial k}\right|_{k=0} = -2\sigma^2 \boldsymbol{N_D}(0)^2 < 0$. Since $\sigma^2 > 0$ and $\boldsymbol{N_D}(0)^2$ is non zero, we conclude that the $MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k), \boldsymbol{b}]$ has decreasing trend at $k = 0$. This implies that there exists at least $k^* > 0$, satisfying $MSE(\hat{\boldsymbol{\beta}}_{\mathbf{RD}}, \boldsymbol{b}) - MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k^*), \boldsymbol{b}] > 0$. Therefore, we can select a suitable positive number k to let the estimator $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k^*)$ performs better than $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$ in the sense of MSE. □

5. MSE-superiority of the difference-based ridge estimator $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(\boldsymbol{k})$ over the differencing estimator $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$

In this section, we provide necessary and sufficient conditions for which the estimator $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$ performs better than $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$ in the sense of $MSE(\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k), \boldsymbol{b}) \leq MSE(\hat{\boldsymbol{\beta}}_{\mathbf{RD}}, \boldsymbol{b})$. From (18) and (19), the difference $\boldsymbol{\Delta} = MSE(\hat{\boldsymbol{\beta}}_{\mathbf{RD}}, \boldsymbol{b}) - MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k), \boldsymbol{b}]$ is given by

$$\boldsymbol{\Delta} = \sigma^2 \boldsymbol{N_D}(0) - \sigma^2 \boldsymbol{N_D}(k) + k\sigma^2 \boldsymbol{N_D}(k)^2 - k^2 \boldsymbol{N_D}(k) \boldsymbol{b}\boldsymbol{b}' \boldsymbol{N_D}(k), \qquad (20)$$

when $\boldsymbol{R}\boldsymbol{b} = \boldsymbol{r}$.

Theorem 5.1. *Let the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{RD}}(k)$ given by under the linear regression model with true restrictions $\boldsymbol{R}\boldsymbol{b} = \boldsymbol{r}$. If $k > 0$, then the MSE difference $\boldsymbol{\Delta}$ is nonnegative definite if and only if*

$$k[\sigma^{-2} \boldsymbol{b}\boldsymbol{b}' - (\boldsymbol{P}\boldsymbol{C_D}\boldsymbol{P})^+] \leq 2\boldsymbol{P}, \qquad (21)$$

where $\boldsymbol{P} = \boldsymbol{I}_p - \boldsymbol{R}'(\boldsymbol{R}\boldsymbol{R}')^{-1}\boldsymbol{R}$. Note that by a + superscript we denote the unique MoorePenrose inverse of a matrix.

Proof. The proof is followed from the proof of Theorem 4.1 of Zhong and Yang (2007) under differencing method. □

6. Choice of the Biasing Parameter

It is difficult to give a satisfying answer about how to select k. This is because the best k always depends on the unknown $\boldsymbol{b}$ and σ^2 in the practical applications which make the problem to be complicated.

Belsley et al. (1980) proposed that the multicollinearity would take effect apparently as the condition number of $\boldsymbol{C_D}$ is bigger than 10. The correlation of the

variables of $\boldsymbol{X}$ is strong comparatively when the condition number of $\boldsymbol{C_D}$ is between 30 and 100. While the condition number is bigger than 100, the correlation would become very strong and the estimated coefficients is very unstable. They suggested that we can obtain the good and stable coefficients if the condition number of $\boldsymbol{C_D}$ is not bigger than 10. Based on this criterion, we can easily choose the k such that the condition number of $\boldsymbol{C_D}$ is reduced to 10 (see Liu, 2003).

Although the criterion mentioned above is simple, our problem to select k is not yet completely solved. Therefore, we give a range to select k in Theorem 4.1.

Theorem 6.1. *Let us be given the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{RD}}(k)$ under the linear regression model with true restrictions $\boldsymbol{Rb} = \boldsymbol{r}$ and $\boldsymbol{b} \neq \boldsymbol{b}_0$. If*

$$0 < k \leq \frac{2\sigma^2}{\boldsymbol{b}'\boldsymbol{P}\boldsymbol{b}}, \tag{22}$$

then the MSE difference $\boldsymbol{\Delta}$ is nonnegative definite.

Proof. The proof is a direct consequence of the proof of Theorem 4 of GroB (2003) under differencing setting. □

Remark 6.1. In the above Theorem if σ^2 is unknown, for driving a range for k which the MSE difference $\boldsymbol{\Delta}$ is nonnegative definite, we estimate σ^2 by

$$\sigma^2 = \frac{1}{n-(p-q)}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathbf{RD}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathbf{RD}}). \tag{23}$$

7. Numerical Study

In this section, we proceed to the comparison of the proposed estimators by some numerical computations. In the scalar comparison, the trace of MSE (see Akdeniz and Erol, 2003; Belsley et al., 1980; Hoerl and Kennard, 1970) will be used as a measurement. That is, we will compare the trace of $MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)]$ and $MSE(\hat{\boldsymbol{\beta}}_{\mathbf{RD}})$ and define scalar Δ as

$$\Delta = \sigma^2 tr[\boldsymbol{N_D}(0)] - \sigma^2 tr[\boldsymbol{N_D}(k)] + k\sigma^2 tr[\boldsymbol{N_D}(k)^2] - k^2 \boldsymbol{b}'\boldsymbol{N_D}(k)^2\boldsymbol{b}. \tag{24}$$

Our sampling experiment consists of different combinations of k. In this study we simulate the response for $n = 5000$ from the following model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{f}(\boldsymbol{t}) + \boldsymbol{\epsilon}, \tag{25}$$

where $\boldsymbol{b} = (-1.5, -2, -3, 5, -4)$, $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ which $\sigma^2 = 4$,

$$f(t_i) = \sqrt{t_i(1-t_i)}\sin\Big(\frac{2.1\pi}{t_i + 0.05}\Big),$$

that is called the Doppler function for $t_i = i/n$, and for $i = 1, ..., n$, $\boldsymbol{x}_i \sim N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ with

$$\boldsymbol{\mu_x} = \begin{pmatrix} 2.5 \\ 2 \\ 3 \\ 1 \\ -1 \end{pmatrix}, \ \boldsymbol{\Sigma}_x = \begin{pmatrix} 1.9 & 1.8 & 1.8 & 1 & 1 \\ 1.8 & 1.8 & 1.8 & 1 & 1 \\ 1.8 & 1.8 & 4.25 & 1 & 1 \\ 1 & 1 & 1 & 2.49 & 1 \\ 1 & 1 & 1 & 1 & 2.25 \end{pmatrix}.$$

In model (6) the parametric effect, $\boldsymbol{b}$, is estimated by a differencing procedure. Optimal differencing weights do not have analytic expressions but may be calculated easily using an optimization routine. Hall et.al. (1990) present weights to order $m = 10$. These contain some minor errors. We use a fourth-order differencing coefficients, $d_0 = 0.8873$, $d_1 = -0.3099$, $d_2 = -0.2464$, $d_3 = -0.1901$, and $d_4 = -0.1409$ in which case $m = 4$. Now we define the $(5000 - 4) \times 5000$ differencing matrix and $\boldsymbol{R}$ respectively given by

$$\boldsymbol{D} = \begin{pmatrix} d_0 & d_1 & d_2 & d_3 & d_4 & 0 & 0 & \dots & 0 \\ 0 & d_0 & d_1 & d_2 & d_3 & d_4 & 0 & \dots & 0 \\ \vdots & \ddots & & & & & & \vdots & \\ 0 & 0 & \dots & 0 & d_0 & d_1 & d_2 & d_3 & d_4 \end{pmatrix} \quad \boldsymbol{R} = \begin{pmatrix} 1 & -2 & 1 & 4 & 5 \\ -1 & 2 & -1 & -3 & 0 \\ 1 & 2 & -1 & -2 & 3 \\ 2 & -1 & 3 & -2 & 0 \end{pmatrix}.$$

k Coefficients	0	1	2	3	4	5
$\hat{\beta}_1$	-1.499217	-1.499220	-1.499223	-1.499226	-1.499229	-1.499233
$\hat{\beta}_2$	-2.008458	-2.008424	-2.008390	-2.008356	-2.008321	-2.008287
$\hat{\beta}_3$	-3.005952	-3.005928	-3.005904	-3.005880	-3.005856	-3.005832
$\hat{\beta}_4$	4.996084	4.996100	4.996116	4.996132	4.996147	4.996163
$\hat{\beta}_5$	-3.999217	-3.999220	-3.999223	-3.999226	-3.999229	-3.999233
$b'(\hat{\boldsymbol{\beta}})b(\hat{\boldsymbol{\beta}})$	0	1.979e-09	7.918e-09	1.781e-08	3.166e-08	4.947e-08
$tr(MSE)$	0.0001915	0.00019155	0.00019154	0.00019153	0.00019153	0.00019153
Δ	0	1.614e-08	2.789e-08	3.523e-08	3.818e-08	3.673e-08
k Coefficients	6	7	8	9	10	11
$\hat{\beta}_1$	-1.499236	-1.499239	-1.499242	-1.499245	-1.499249	-1.499252
$\hat{\beta}_2$	-2.008253	-2.008219	-2.008185	-2.008150	-2.008116	-2.008082
$\hat{\beta}_3$	-3.005808	-3.005784	-3.005759	-3.005735	-3.005711	-3.005687
$\hat{\beta}_4$	4.996179	4.996195	4.996211	4.996227	4.996243	4.996258
$\hat{\beta}_5$	-3.999236	-3.999239	-3.999242	-3.999245	-3.999249	-3.999252
$b'(\hat{\boldsymbol{\beta}})b(\hat{\boldsymbol{\beta}})$	7.123e-08	9.695e-08	1.266e-07	1.602e-07	1.978e-07	2.393e-07
$tr(MSE)$	0.00019153	0.00019154	0.00019156	0.00019158	0.00019160	0.00019163
Δ	3.088e-08	2.063e-08	5.993e-09	-1.304e-08	-3.646e-08	-5.428e-08

All computations were conducted using the R statistical package. The matrix $\boldsymbol{C_D}$ has eigenvalues: $\lambda_1 = 237.04$, $\lambda_2 = 5723.24$, $\lambda_3 = 6684.99$, $\lambda_4 = 11883.56$ and $\lambda_5 = 39604.97$. Thus the ratio of the largest eigenvalue to the smallest eigenvalue is $\lambda_5/\lambda_1 = 167.07$ which implies the existence of multicollinearity in the data set.

We estimated σ^2 by making use of the equation (23) which are equal to 4.2683 in the model (25).

Table 1 gives several results including $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$, $b'[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)]b[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)]$,

$tr(MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)])$ and Δ when k is in different values in model (25). From this Table, the Δ increases ($tr(MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)])$ decreases) at first and then decreases (increases). Furthermore, the maximum of Δ (minimum of the $tr(MSE[\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)])$) is obtained when k equals to median range of (22) i.e., $\frac{\sigma^2}{\boldsymbol{b}'\boldsymbol{P}\boldsymbol{b}}$ which is approximately equal to 4.17 in models (25), as it can be seen in Figure 1.

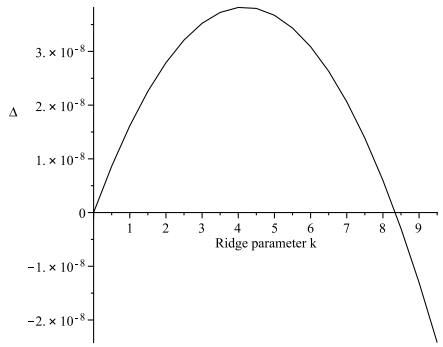

Fig. 1. The diagram of Δ versus k.

In Figure 2, the nonparametric part of the model (25) is plotted, in the top left plot. This functions are difficult to estimate and provide a good test case for nonparametric regression methods. The function is spatially inhomogeneous which means that it's smoothness (second derivatives) varies over x. The top right plot shows $n = 5000$ data points after removing the linear part, i.e., $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$. The middle left and right plot shows the residuals which obtained after estimation of the linear part of the models by $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$, i.e., $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$ and the fitted function, respectively. The bottom left and right plot is the middle part when $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$ replaced with $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(4)$.

According to Table 1 and Figure 1, it can be realized that for all combinations of k, $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$ is better than $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$ if $k \leq a$, which $a = \frac{2\sigma^2}{\boldsymbol{b}'\boldsymbol{P}\boldsymbol{b}}$ and it is equal to 8.339113 in model (25). Furthermore, as it can be realized from Figures 2, the estimators of wavy function is reasonable using the discussed method in this approach.

References

1. Ahn, H. and Powell, J. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *J. Econom.*, **58**, 3-29.
2. Akdeniz, F. and Erol, H. (2003). Mean squared error matrix comparisons of some biased estimators in linear regression. *Commun. Statist. Theory Methods*, **32**, 23892413.
3. Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics.* Wiley, New York.

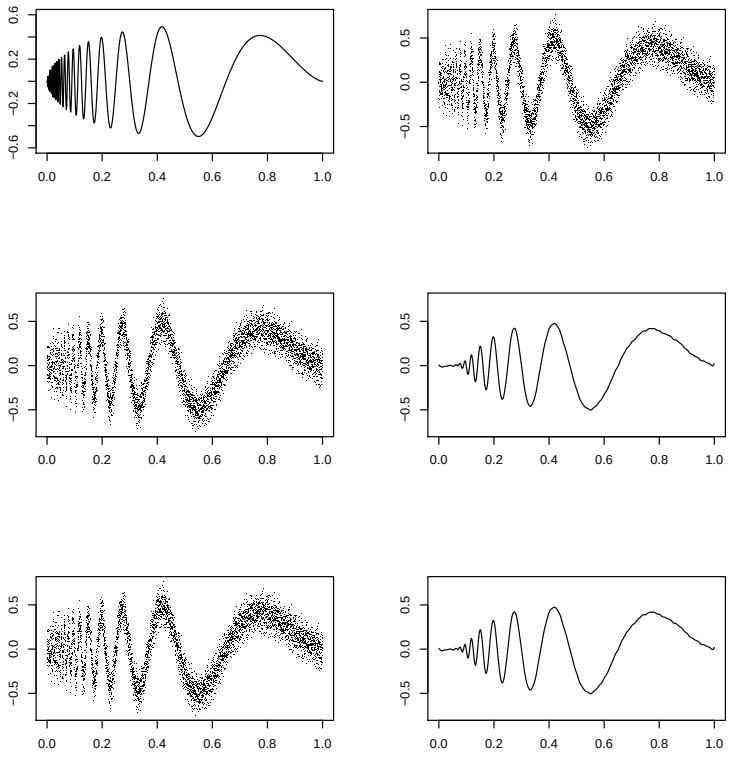

Fig. 2. Estimation of the $f(\cdot)$ (Doppler function) by local linear regression for $n = 5000$. The function (top left), the data (top right), the residuals obtained after estimation of the linear part of the model by $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}$ (middle left), the fitted function for $k = 0$ (middle right), the residuals obtained after estimation of the linear part of the model by $\hat{\boldsymbol{\beta}}_{\mathbf{RD}}(k)$ for $k = 4$ (bottom left), the fitted function for $k = 4$ (bottom right).

4. Grob, J., (2003). Restricted ridge estimation. *Statist. Prob. Lett.*, **65**, 57-64.
5. Hall, P. Kay, J.W. and Titterington, D.M., (1990). On estimation of noise variance in two-dimentional signal processing. *Advanced in appl. Prob.* , **23**, 476-495.
6. Hoerl, A.E. and Kennard, R.W., (1970). Ridge regression biased estimation for non-orthogonal problems. *Technometrics* , **12**, 69-82.
7. Liu, K. J. (2003). Using Liu type estimator to combat collinearity. *Commun. Statist. Theo. Meth.*, **32**, 1009-1020.
8. Roozbeh, M., Arashi, M. and Niroumand H. A. (2009). Ridge Regression Methodology in Partial Linear Models with Correlated Errors.*Statist. Comput. Simu.*, (in press).
9. Sarkar, N., (1992). A new estimator combining the ridge regression and the

restricted least squares methods of estimation. *Comm. Statist. Theo. Meth.*, **21**, 1987-2000.

10. Swamy, P. A. V. B. and Mehta, J. S. (1977). A note on minimum average risk estimator for coefficient in linear models. *Comm. Statist.* , **6**, 1161-1186.
11. Swamy, P. A. V. B., Mehta, J. S. and Rappoport, P. N. (1978). Two methods of evaluating Hoerl and Kennard's ridge regression. *Comm. Statist.*, **12**, 1133-1155.
12. Wang, L., Brown, L.D. and Cai, T.T., (2007). A difference based approach to semiparametric partial linear model. Department of Statistics, The Wharton School University of Pennsylvania, Technical report.
13. Yatchew, A., (1997). An elemantary estimator of the partial linear model. *Econ. Lett.*, **57**, 135-143.
14. Yatchew, A., (2000). Scale economies in electricity distribution: a semiparametric analysis. *J. Appl. Econom.*, **57**, 187-210.
15. Yatchew, A., (2003). *Semiparametric regression for the applied econometrican.* Cambridge University press, Cambridge.
16. Zhong, Z. and Yang, H. (2007). Ridge Estimation to the Restricted Linear Model. *Comm. Statist. Theo. Meth.*, **36**, 2099-2115.

Results on the Past Lifetime of $(n-k+1)$-out-of-n Structures with Nonidentical Components

E. Salehi and M. Asadi

Department of Statistics, University of Isfahan

We consider a $(n-k+1)$-out-of-n system with independent and nonidentical components. Under the condition that at time t the system has failed we study the past lifetime of the components of the system. The mean past lifetime of the components is defined and some of its properties are investigated. Stochastic comparisons are also made between the past lifetime of different systems.

Keywords: Order statistics; Reversed hazard rate; Stochastic ordering; Reliability.

1. Introduction

The $(n-k+1)$-out-of-n systems are very popular type of redundancy in technical systems and structures. A $(n-k+1)$-out-of-n system is a system consisting of n components and the system functions if and only if at least $(n-k+1)$ components out of n components are operating $(k \leq n)$. Two important special cases of $(n-k+1)$-out-of-n systems are parallel systems and series systems corresponding to $k=n$ and $k=1$, respectively. In recent years several authors have studied the reliability and aging properties of $(n-k+1)$-out-of-n systems. Among others, we refer to Asadi and Bayramoglu (2006), Asadi (2006), Khaledi and Shaked (2007), Li and Zhao (2008), and Navarro et al. (2008).

Consider a $(n-k+1)$-out-of-n system and assume that $X_1, X_2, ..., X_n$ denote the lifetimes of the components of the system. Denote by $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ the ordered lifetimes. Suppose that the system, which starts to work at $t=0$, is not working at time $t>0$. On the basis of structure of the system if the failure times of the components are not monitored continuously then the exact failure times of components with lifetimes $X_{1:n}, X_{2:n}, \ldots, X_{k:n}$ are unknown. Hence it might be important for engineers and system designers to have some information about the average time elapsed since the failure of the components. The time that has elapsed from the failure of $X_{l:n}, l=1,2,\ldots,k$, given that system has failed at or before t is $X_t^{l,k,n} = t - X_{l:n} | X_{k:n} \leq t$.

In recent years some authors studied the properties of the conditional random variables of the form $X_t^{l,k,n}$. Most of the results in the literature are dealing with the case in which the system has independent and identical components, i.e the X_i's are independent and have the same distribution function F. Under these conditions Asadi (2006) considered the past lifetime of the components of a parallel system and defined the concept of mean past lifetime (MPL) of the components as follows $M_n^k(t) = E(t - X_{k:n}|X_{n:n} \leq t)$, $k = 1, 2, \ldots, n$. Tavangar and Asadi (2009) extended the result of Asadi (2006) to $(n-k+1)$-out-of-n system and defined the MPL of the components of the system as follows $M_n^{l,k}(t) = E\left(t - X_{l:n} \mid X_{k:n} \leq t\right)$, $1 \leq l \leq k \leq n$. Khaledi and Shaked (2007) studied the stochastic properties of conditional random variables of forms $X_t^{l,k,n}$ in coherent systems. In real life situation the systems might be designed with nonidentical components. Motivated by this, Sadegh (2008) extended the results of Asadi (2006) for the parallel systems with nonidentical components.

The aim of the present paper is to extend the results of Sadegh to $(n-k+1)$-out-of-n systems and study the properties of the MPL $M_n^{l,k}(t)$ with nonidentical components. When the independent random variables are not identically distributed, the distributions of order statistics can be expressed in terms of permanents. Let S_n be the set of permutations of $\{1, 2, \ldots, n\}$. For an $n \times n$ matrix, say $A = (a_{i,j})$, the permanent of A denoted by $PerA$, is defined as

$$PerA = \sum_{\pi \in S_n} \prod_{i=1}^{n} a_{i,\pi(i)},$$

where $\pi = (\pi(1), \pi(2), \ldots, \pi(n))$. If $a_1, a_2, \ldots,$ denote the column vectors, then $[\underbrace{a_1}_{i_1}, \underbrace{a_2}_{i_2}, \ldots,] = [a_1\mathbf{1}'_{i_1}, a_2\mathbf{1}'_{i_2}, \ldots,]$, where $\mathbf{1}_{i_1}$ shows a column vector of 1s of length i_1. The concept of permanent is like the concept of determinant, except that all signs in the expansion are positive.

When $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ are order statistics of a set of independent random variables with absolutely continuous distribution functions $F_1, F_2, \ldots, F_n$ and densities $f_1, f_2, \ldots, f_n$, respectively, Vaughan and Venables (1972) showed that the density of $X_{r:n}$ can be expressed in terms of permanents. Bapat and Beg (1989) showed that the distribution function of $X_{r:n}$, $(1 \leq r \leq n)$ is given by

$$P(X_{r:n} \leq x) = \sum_{i=r}^{n} \frac{1}{i!(n-i)!} Per \left[\underbrace{\mathbf{F}(\mathbf{x})}_{i}, \underbrace{\bar{\mathbf{F}}(\mathbf{x})}_{n-i}\right], \qquad -\infty < x < \infty,$$

where $\mathbf{F}(\mathbf{x}) = (F_1(x), \ldots, F_n(x))'$, $\bar{\mathbf{F}}(\mathbf{x}) = (\bar{F}_1(x), \ldots, \bar{F}_n(x))$.

This paper is organized as follows. In Section 2 we first obtain the survival function of $X_t^{l,k,n}$. We show that the random variable $X_t^{l,k,n}$ is stochastically ordered in terms of l and k, which in turn implies that $M_n^{l,k}(t)$ is ordered in terms of l and k. We consider, in this section, two systems with same number of components and show that when the components of the systems are ordered in terms of reversed hazard rates then the corresponding MPL of the systems are also ordered. In Section 3 we study the aging properties of the system. It is shown that when a component has a bathtub-shaped reversed hazard then the MPL of the component has an upsidedown bathtub-shaped form. It is shown by an example that the result is not necessary true for the system. In other words, we have shown when the components of the system have bathtub-shaped reversed hazard rates then the MPL $M_n^{l,k}(t)$ may not have an upsidedown bathtub-shaped form for all values of l and k.

2. The Past Lifetime of the Components of the Systems

In this section we assume that a $(n-k+1)$-out-of-n system consists of n independent components with lifetimes $X_1, X_2, \ldots, X_n$ where X_i is distributed as F_i, $i = 1, 2, \ldots, n$. We denote the survival function of X_i by $\bar{F}_i$. We obtain several properties of $X_t^{l,k,n}$, the past lifetime of the components of the system. If we denote the survival function of $X_t^{l,k,n}$ by $\bar{F}_{l,k,n,t}$ then we can prove the following theorem using combinatorial arguments.

Theorem 2.1. *For $1 \leq l \leq k \leq n$ and $0 < x < t$,*

$$\bar{F}_{l,k,n,t}(x) = P(X_t^{l,k,n} > x) = \frac{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t) \bar{F}^{(C_i)}_{n-i-l+1,n-i,t}(x)}{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t)}, \qquad (1)$$

where C_i denotes a subset of $\{1, 2, \ldots, n\}$ with i elements, C_i^c is the complement of C_i, $C_0 = \phi$, $\psi_i(t) = \prod_{s \epsilon C_i} \frac{\bar{F}_s(t)}{F_s(t)}$ and

$$\bar{F}^{(C_i)}_{n-i-l+1,n-i,t}(x) = \sum_{j=0}^{n-i-l} \frac{1}{j!(n-i-j)!} Per \left[\underbrace{\bar{\mathbf{F}}_{\mathbf{t}}(\mathbf{x})}_{j} \ , \ \underbrace{\mathbf{F}_{\mathbf{t}}(\mathbf{x})}_{n-i-j} \right]_{C_i^c},$$

where $\mathbf{F_t(x)}$ is a vector in which the ith element is $\frac{F_i(t-x)}{F_i(t)}$.

Remark 2.1. If $k = n$, then $\bar{F}_{l,k,n,t}(x)$ denotes the survival function of past lifetime of the components of a parallel system with independent and non-identical components which is already derived by Sadegh (2008). If X_i are i.i.d random variables, that is, the system has independent and identical components the representation (1) reduces to the result of Tavangar and Asadi (2009).

We denote the expectation of $X_t^{l,k,n}$ by $M_n^{l,k}(t)$. That is $M_n^{l,k}(t) = E(t - X_{l:n}|X_{k:n} \leq t)$. $M_n^{l,k}(t)$ is called the MPL (or mean inactivity) of the system components at the system level. For $l = 1, 2, \ldots, k$ the MPL $M_n^{l,k}(t)$ can be written as

$$M_n^{l,k}(t) = \frac{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t) M_{n-i,C_i}^{n-i-l+1}(t)}{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t)}, \tag{2}$$

where

$$M_{n-i,C_i}^{n-i-l+1}(t) = E(t - X_{n-i-l+1:n-i}^{(C_i)}|X_{n-i:n-i}^{(C_i)} \leq t),$$

is the MPL of $X_{n-i-l+1:n-i}^{(C_i)}$, $l = 1, 2, \ldots, n-i$, in a parallel system consisting of $n-i$ non-identical components.

Remark 2.2. Note that MPL $M_n^{l,k}$ can also be represented as follows:

$$M_n^{l,k}(t) = \sum_{i=0}^{n-k} \sum_{C_i} p^{(C_i)}(t) M_{n-i,C_i}^{n-i-l+1}(t),$$

where

$$p^{(C_i)}(t) = \frac{\psi_i(t)}{\sum_{j=0}^{n-k} \sum_{C_j} \psi_j(t)} = P\left(Z_t^{C_i} = n - |C_i| \Big| \; Z_t \geq k\right),$$

and $Z_t = \sum_{i=1}^{n} X_{t,i}$ and $X_{t,i}$ $i = 1, 2, \ldots, n$ are n independent random variables distributed as binomial with parameters $(1, F_i(t))$. This shows that the MPL of $X_{l:n}$, $l \leq k$, of $(n-k+1)$-out-of-n system can be represented as weighted mean of the MPL's of $X_{n-i-l+1:n-i}$, $l = 1, 2, \ldots, n-i$, of parallel system.

Now, we have the following results.

Theorem 2.2. *For $1 \leq l \leq k < n$, and $t > 0$,*

$$X_t^{l,k,n} \leq_{st} X_t^{l,k+1,n},$$

where st stands for the usual stochastic order (for more details see Shaked and Shantikumar (2007)).

Proof. Note that, for $1 \leq l \leq k < n$ and $0 < x < t$,

$$\begin{aligned} &\bar{F}_{l,k,n,t}(x) - \bar{F}_{l,k+1,n,t}(x) \\ &= \frac{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t) \bar{F}_{n-i-l+1,n-i,t}^{(C_i)}(x)}{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t)} - \frac{\sum_{i=0}^{n-k-1} \sum_{C_i} \psi_i(t) \bar{F}_{n-i-l+1,n-i,t}^{(C_i)}(x)}{\sum_{i=0}^{n-k-1} \sum_{C_i} \psi_i(t)}. \end{aligned}$$

We need to show that the right hand side of this equation is non-positive. To this end, it is enough to show that, for $0 \leq i \leq n-k-1$, the following quantity is non-positive

$$\sum_{i=0}^{n-k-1} \sum_{C_i} \sum_{C_{n-k}} \prod_{s \epsilon C_i + C_{n-k}} \frac{\bar{F}_s(t)}{F_s(t)} \bar{F}^{(C_{n-k})}_{k-l+1,k,t}(x) - \sum_{i=0}^{n-k-1} \sum_{C'_i} \sum_{C'_{n-k}} \prod_{s \epsilon C'_i + C'_{n-k}} \frac{\bar{F}_s(t)}{F_s(t)} \bar{F}^{(C'_i)}_{n-i-l+1,n-i,t}(x). \quad (3)$$

Both expressions in (3) have the same number of terms. For each C_i, we can choose C'_{n-k} such that $C_i \subset C'_{n-k}$ and $C'_i + C'_{n-k} = C_i + C_{n-k}$. Thus, (3) can be represented as follows

$$\sum_{i=0}^{n-k-1} \sum_{C_i \subset C'_{n-k}} \prod_{s \epsilon C_i + C_{n-k}} \frac{\bar{F}_s(t)}{F_s(t)} \left[\bar{F}^{(C'_{n-k})}_{k-l+1,k,t}(x) - \bar{F}^{(C_i)}_{n-i-l+1,n-i,t}(x) \right].$$

Since $C'^c_{n-k} \subset C^c_i$, using Lemma Nanda and Shaked (2001), we have $\bar{F}^{(C'_{n-k})}_{k-l+1,k,t}(x) \leq \bar{F}^{(C_i)}_{n-i-l+1,n-i,t}(x)$ which gives the required result. □

Theorem 2.3. *For $1 \leq l < k \leq n$, and $t > 0$,*

$$X^{l+1,k,n}_t \leq_{st} X^{l,k,n}_t.$$

Proof. Note that, for $1 \leq l < k \leq n$, $0 < x < t$,

$$\bar{F}_{l+1,k,n,t}(x) - \bar{F}_{l,k,n,t}(x) = \frac{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t) \bar{F}^{(C_i)}_{n-i-l,n-i,t}(x)}{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t)} - \frac{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t) \bar{F}^{(C_i)}_{n-i-l+1,n-i,t}(x)}{\sum_{i=0}^{n-k} \sum_{C_i} \psi_i(t)}.$$

We show that the right hand side of the above equation is non-positive. To show this, it is enough to show that the following equation is non-positive.

$$\sum_{i=0}^{n-k} \sum_{j=0}^{n-k} \sum_{C_i} \sum_{C_j} \psi_i(t) \psi_j(t) \left[\bar{F}^{(C_i)}_{n-i-l,n-i,t}(x) - \bar{F}^{(C_i)}_{n-i-l+1,n-i,t}(x) \right].$$

Now from Lemma Nanda and Shaked (2001), we have $\bar{F}^{(C_i)}_{n-i-l,n-i,t}(x) \leq \bar{F}^{(C_i)}_{n-i-l+1,n-i,t}(x)$ for all $i = 0, 1, \ldots, n-k$, which gives the required result. □

Corollary 2.1.

(a) Under the conditions of Theorem 2.2,,for fixed values of l and n, $M^{l,k}_n(t)$ is a increasing function of k, $k = l, 2, \ldots, n$;

(b) Under the conditions of Theorem 2.3,for fixed values of k and n, $M^{l,k}_n(t)$ is a decreasing function of l, $l = 1, 2, \ldots, k$.

Khaledi and Shaked (2007) proved that when the components of two $(n-k+1)$-out-of-n systems φ_1 and φ_2(each consisting of n independent and identical components, where components of φ_1 and φ_2 are distributed according to function F and G, respectively) are ordered in term reversed hazard rates, then the systems are ordered in terms of their MPL's. In other words, they showed that if $r_F(t) \leq r_G(t),\ t > 0$, (reversed hazard of absolutely distribution F with density function f is $r_F(t) = \frac{f(t)}{F(t)}$) then $M_n^{l,k}(t) \geq K_n^{l,k}(t)$, where $M_n^{l,k}(t)$, is MPL of φ_1 and $K_n^{l,k}(t)$, is MPL of φ_2. The following theorem extends the result of Khaledi and Shaked (2007).

Theorem 2.4. *Let $X_1, X_2, \ldots, X_n$, $Y_1, Y_2, \ldots, Y_n$ be independent continuous random variables. If, either X_i or Y_i are identically distributed, and for any $1 \leq i, j \leq n$, $X_i \leq_{rh} Y_j$, where rh stands for reversed hazard rate order, then, for $1 \leq l \leq k \leq n$, and $t > 0$,*

$$X_t^{l,k,n} \geq_{st} Y_t^{l,k,n}.$$

Proof. We give the proof for the case where $Y_i, i = 1, 2, \ldots, n$ are identically distributed. Similar proof can be given for the case when $X_i, i = 1, 2, \ldots, n$ are identically distributed. To get the result, it is enough to prove that $\bar{G}_{l,k,n,t}(x) - \bar{F}_{l,k,n,t}(x) \leq 0$, or to show that the following expression is non-positive.

$$\sum_{i=0}^{n-k} \sum_{C_i} \psi_i^Y(t) \bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x) \sum_{i=0}^{n-k} \sum_{C_i} \psi_i^X(t) - \sum_{i=0}^{n-k} \sum_{C_i} \psi_i^X(t) \bar{F}^{(C_i)}_{n-i-l+1,n-i,t}(x) \sum_{i=0}^{n-k} \sum_{C_i} \psi_i^Y(t), \quad (4)$$

where $\psi_i^Y(t) = \prod_{s \epsilon C_i} \frac{\bar{G}_s(t)}{G_s(t)}$, $\psi_i^X(t) = \prod_{s \epsilon C_i} \frac{\bar{F}_s(t)}{F_s(t)}$. Note that for $0 \leq x \leq t$, (4) is equal

$$\sum_{i=0}^{n-k} \sum_{j=0}^{n-k} \sum_{C_i} \sum_{C_j} \psi_j^X(t) \psi_i^Y(t) \left[\bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x) - \bar{F}^{(C_j)}_{n-j-l+1,n-j,t}(x) \right]$$

$$= \sum_{i=0}^{n-k} \sum_{j=0}^{n-k} \sum_{C_i} \sum_{C_j} \psi_j^X(t) \psi_i^Y(t) \Big[\bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x) - \bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x) + \bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x) - \bar{F}^{(C_j)}_{n-j-l+1,n-j,t}(x) \Big]$$

$$= \sum_{i=0}^{n-k}\sum_{j=0}^{n-k}\sum_{C_i}\sum_{C_j}\psi_j^X(t)\psi_i^Y(t)\left[\bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x) - \bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x)\right]$$
$$+ \sum_{i=0}^{n-k}\sum_{j=0}^{n-k}\sum_{C_i}\sum_{C_j}\psi_j^X(t)\psi_i^Y(t)\left[\bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x) - \bar{F}^{(C_j)}_{n-j-l+1,n-j,t}(x)\right]$$
$$= \phi_1 + \phi_2.$$

We will show that both ϕ_1 and ϕ_2 are non-positive. We have

$$\phi_1 = \sum_{i=0}^{n-k}\sum_{j=i}^{n-k}\sum_{C_i}\sum_{C_j}\psi_j^X(t)\psi_i^Y(t)\left[\bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x) - \bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x)\right]$$
$$+ \sum_{i=0}^{n-k}\sum_{j=0}^{i}\sum_{C_i}\sum_{C_j}\psi_j^X(t)\psi_i^Y(t)\left[\bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x) - \bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x)\right]$$
$$= \sum_{i=0}^{n-k}\sum_{j=i}^{n-k}\sum_{C_i}\sum_{C_j}\psi_j^X(t)\psi_i^Y(t)\left[\bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x) - \bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x)\right]$$
$$+ \sum_{i=0}^{n-k}\sum_{j=i}^{n-k}\sum_{C_i}\sum_{C_j}\psi_i^X(t)\psi_j^Y(t)\left[\bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x) - \bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x)\right]$$
$$= \sum_{i=0}^{n-k}\sum_{j=i}^{n-k}\sum_{C_i}\sum_{C_j}\left[\psi_j^X(t)\psi_i^Y(t) - \psi_i^X(t)\psi_j^Y(t)\right]$$
$$\left[\bar{G}^{(C_i)}_{n-i-l+1,n-i,t}(x) - \bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x)\right].$$

Under the assumption that Y_i's are i.i.d random variables, the above equation can be rewritten as

$$\sum_{i=0}^{n-k}\sum_{j=0}^{n-k}\sum_{C_i}\sum_{C_j}\left[\prod_{s\epsilon C_j}\frac{\bar{F}_s(t)}{F_s(t)}\left(\frac{\bar{G}(t)}{G(t)}\right)^i - \prod_{s\epsilon C_i}\frac{\bar{F}_s(t)}{F_s(t)}\left(\frac{\bar{G}(t)}{G(t)}\right)^j\right]$$
$$\times\left[\bar{G}_{n-i-l+1,n-i,t}(x) - \bar{G}_{n-j-l+1,n-j,t}(x)\right]. \quad (5)$$

By using Lemma Nanda and Shaked (2001), we have

$$\left[\bar{G}_{n-i-l+1,n-i,t}(x) - \bar{G}_{n-j-l+1,n-j,t}(x)\right] \geq 0,$$

for $0 \leq i \leq j \leq n-k$. Hence to show that ϕ_1 is non-positive, it is enough to show the first expression in summations in (5) is non-positive. But for $i \leq j$, we have

$$\sum_{C_i}\sum_{C_j}\left[\prod_{s\epsilon C_j}\frac{\bar{F}_s(t)}{F_s(t)}\left(\frac{\bar{G}(t)}{G(t)}\right)^i - \prod_{s\epsilon C_i}\frac{\bar{F}_s(t)}{F_s(t)}\left(\frac{\bar{G}(t)}{G(t)}\right)^j\right]$$
$$= \sum_{C_j}\prod_{s\epsilon C_j}\frac{\bar{F}_s(t)}{F_s(t)}\sum_{C_i}\left(\frac{\bar{G}(t)}{G(t)}\right)^i - \sum_{C_i}\prod_{s\epsilon C_i}\frac{\bar{F}_s(t)}{F_s(t)}\sum_{C_j}\left(\frac{\bar{G}(t)}{G(t)}\right)^j.$$

Since the first part of above expression has same number of terms as the second part, for each C_i, we can choose C_j, such that $C_i \subset C_j$. Hence, the above equation could be represented as

$$\sum \sum_{C_i \subset C_j} \prod_{s \epsilon C_i} \frac{\bar{F}_s(t)}{F_s(t)} \left(\frac{\bar{G}(t)}{G(t)}\right)^i \left[\prod_{s \epsilon C_{j-i}} \frac{\bar{F}_s(t)}{F_s(t)} - \left(\frac{\bar{G}(t)}{G(t)}\right)^{(j-i)}\right].$$

From the fact that $X \leq_{rh} Y$ one can easily see that this last expression is non-positive. On the other hand ϕ_2 can be written as

$$\phi_2 = \sum_{i=0}^{n-k} \sum_{j=0}^{n-k} \sum_{C_i} \sum_{C_j} \psi_j^X(t) \psi_i^Y(t) \left[\bar{G}^{(C_j)}_{n-j-l+1,n-j,t}(x) - \bar{F}^{(C_j)}_{n-j-l+1,n-j,t}(x)\right].$$

The assumption that $X \leq_{rh} Y$ implies easily that $(t-Y|Y<t) \leq_{st} (t-X|X<t)$. This in turn, implies that ϕ_2 is non-positive. Hence, the proof is complete. □

Corollary 2.2. *Assume that $M_n^{l,k}(t)$ and $K_n^{l,k}(t)$ denote the MPLs of the systems corresponding to F and G described in above theorem. Then under the assumptions of the theorem we have*

$$M_n^{l,k}(t) \geq K_n^{l,k}(t).$$

3. Some aging properties of the system

Let X be a non-negative absolutely continuous random variable with distribution function F, density function f, survival function $\bar{F} = 1-F$ and hazard rate $h(t) = \frac{f(t)}{\bar{F}(t)}$. Mi (1996) proved that when the distribution function F has a bathtub-shaped hazard rate $h(t)$ then the corresponding mean residual life function $m(t)$ has an upsidedown bathtub-shape. Recall that a hazard rate $h(t)$ is said to be bathtub-shaped if there exist two change points t_1 and t_2 $(0 \leq t_1 \leq t_2 \leq \infty)$ such that $h(t)$ is strictly decreasing on $(0, t_1)$, is constant on (t_1, t_2) and is strictly increasing on (t_2, ∞). In the following theorem we prove that when the reversed hazard rate $r(t)$ is bathtub-shaped then the corresponding MPL $M(t)$ has upsidedown bathtub-shaped.

Theorem 3.1. *Let the distribution function F of a lifetime have a differentiable bathtub shaped reverse hazard $r(t)$ with change points t_1 and t_2 $(0 \leq t_1 \leq t_2 \leq \infty)$. Then there exists a unique time t^* such that $t_2 < t^* \leq \infty$ such that*

$$M(t^*) = \max_{t \geq 0} \frac{\int_0^t F(x)dx}{F(t)}$$

Proof. Differentiating of $M(t)$ in terms of t gives

$$M'(t) = 1 - r(t)M(t). \tag{6}$$

Define $A(t) = F(t) - r(t)\int_0^t F(x)dx$. Hence, $M^{'}(t) = \frac{A(t)}{F(t)}$. The derivative of $A(t)$ is given by $A'(t) = -r'(t)\int_0^t F(x)dx$. Hence, under the assumptions on $r(t)$ we can show that

$$A(t)\begin{cases} \text{is strictly increasing,} & \text{if } 0 \le t \le t_1; \\ \text{is constant,} & \text{if } t_1 < t < t_2; \\ \text{is strictly decreasing,} & \text{if } t_2 \le t. \end{cases}$$

Note that for $t \le t_1$ it can be concluded that

$$\begin{aligned} A(t) &= F(t)\left(1 - r(t)\int_0^t \exp\left\{-\int_x^t r(u)du\right\}dx\right) \\ &> F(t)\exp\left\{-\int_0^t r(u)du\right\} \\ &\ge 0. \end{aligned}$$

Hence $A(t) > 0$, $\forall t \le t_2$ since $A(t)$ is a constant on $[t_1, t_2]$. Thus, from representation (6), $M(t)$ is always strictly increasing in $t \le t_2$. If $A(\infty) \ge 0$, then $A(t) > 0$ and hence for all $t > 0$, $M'(t) \ge 0$ which implies that $M(t)$ is strictly increasing. On the other hand, if $A(\infty) < 0$, then there exits a unique $t^* \in (t_2, \infty]$ such that, $A(t^*) = 0$ and

$$A(t)\begin{cases} > 0, & \text{if } 0 \le t \le t^*; \\ = 0, & \text{if } t = t^*; \\ < 0, & \text{if } t^* \le t. \end{cases}$$

This means that $M(t)$ strictly increases (decreases) for $0 \le t \le t^*$ ($t^* \le t$), that is MPL takes its maximum in $t = t^*$, or $M(t^*) = \max_{t \ge 0} M(t)$. Hence, the proof is complement. □

The following example gives an application of this theorem.

Example 3.1. Let X_1, X_2, X_3 denote the lifetime of three components which are connected in a system. Assume that X_i's are distributed, respectively, as

$$\begin{aligned} f_1(x) &= 3(1-x)^2/2, \quad 0 < x < 2, \\ f_2(x) &= 5(1-x)^4/2 \quad 0 < x < 2, \\ f_3(x) &= 7(1-x)^6/2 \quad 0 < x < 2. \end{aligned}$$

Figures 1 and 2 show the graphs of reversed hazards, and the MPLs of X_i's, respectively. It is seen from the graphs that the reversed hazard rates are bathtub-shaped and the MPLs are upsidedown bathtub-shaped.

A natural question here is whether the result of the above theorem is true for the systems. In other words, can we conclude that when the reversed hazard rate $r(t)$ has bathtub-shaped then the MPL of the system, $M_n^{l,k}(t)$, for all l, k and n, has an upsidedown bathtub-shaped? The answer in general is negative. To see this we consider a system consisting of 3 components and assume that the components of the system are distributed according to Example 3.1. The graphs of $M_n^{l,k}$ for the case where $n = 3$, $k = 2$ and $l = 1, 2$ is given in Figure 3 and the graphs of $M_n^{l,k}$ for the case where $n = 3$, $k = 3$ and $l = 1, 2, 3$ is presented in Figure 4. The graphs in Figure 4 show that when the components have bathtub-shaped reversed hazard rates then the MPL of the system for different values of l, $l = 1, 2, 3$ is upsidedown bathtub-shaped. However, the graphs of Figure 3 indicate that the MPL of the system is not necessarily upsidedown bathtub-shaped. The graph of $M_3^{1,2}(t)$ (at the bottom of Figure 3) is first upsidedown but after sometime starts to increase.

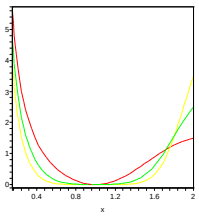

Fig. 1.

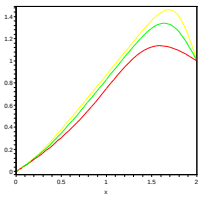

Fig. 2.

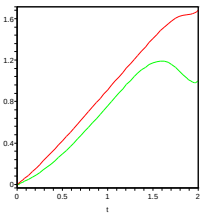

Fig. 3.

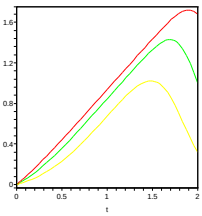

Fig. 4.

References

1. M. Asadi, (2006). On the mean past lifetime of the components of a parallel system. *J. Statist. Plann. Inference*, **136**, 1197-1206.
2. M. Asadi, and I. Bayramoglu, (2006). The mean residual life function of a k-out-of-n structure at the system level. *IEEE Trans. Reliab.*, **55**, 314-318.
3. R. B. Bapat, and M. I. Beg, (1989). Order statistics for nonidentically distributed variables and permanents, *Sankhya*: Indian J. Stat. 51 (Ser. A) 79-93.
4. B.-E. Khaledi, and M. Shaked, (2007). Ordering conditional lifetimes of coherent systems. *J. Statist. Plann. Inference*, **137**, 1173-1184.
5. X. Li, and P. Zhao, (2008). Stochastic comparison on general inactivity time and general residual life of k-out-of-n systems. *Comm. Statist. Simu. Comp.* **37**, 1005-1019.
6. J. Mi, (1996). Minimizing some cost functions related to both burn-in and field use. *Oper. Res.*, **44**, 497-500.
7. A. K. Nanda, and M. Shaked, (2001). The hazard rate the reversed hazard rate orders, with applications to order statistics. *Ann. Inst. Statist. Math.*, **53**, 853-864.
8. J. Navarro, N. Balakrishnan, and F. J. Samaniego, (2008). Mixture representations of residual lifetimes of used systems. *Appl. Probab.*, **45**, 1097-1112.

9. M. K. Sadegh, (2008). Mean past and mean residual Life of parallel System with Nonidentical Components. *Comm. Statist. Theory Methods*, **37**, 1137-1145.
10. M. Shaked, and J. G. Shanthikumar, (2007). *Stochastic orders and their applications*. Springer, New York.
11. M. Tavangar, and M. Asadi, (2009). A study on the mean past lifetime of the components of $(n-k+1)$-out-of-n system at the system level. *To appear in Metrica.*
12. R. J. Vaughan, and W. N. Venables, (1972). Permanent expressions for order statistics densities, *J. Royal Stat. Soc.* B 34 (2) 308-310.

A Modified Rank Test For Model Selection

A. Sayyareh

Statistics Department, Razi University

In this paper we consider a distribution-free test for nonnested distributions in the context of the model selection. We introduced a new test and showed that this test is asymptotically more efficient than the Vuong test and the test statistic based on B statistic introduced by Clarke. This work is heavily based on the Clarke approach to finding a test. But here we let to the magnitude of the data to give a better performance to the test statistic. We showed that this test is an unbiased test.

Keywords: Cox'S Test, Hypothesis Testing, Kullback-Leibler Divergence, Model Selection, Mis-Specified Models, Non-Nested Models, Rank Test, Vuong'S Test.

1. Introduction

As we know rank statistics are distribution-free under the null hypothesis, the level of a rank test is independent of the underlying distribution, which is the best possible protection of the level against misspecification of the model. On the other hand, the power of a rank test is not necessarily robust against deviations from the postulated model. This might lead to the use of the best test for the wrong model. Let $\bar{Y} = (Y_1, Y_2, ..., Y_n)$ be identically and independently distributed random variables from unknown density $h(.)$. Two rival model are assumed as possible explanation of Y, represented by $(f^{\gamma}(.))_{\gamma\in\Gamma} = \{f(y;\gamma), \gamma \in \Gamma\}$ and $(g^{\beta}(.))_{\beta\in B} = \{g(y;\beta), \beta \in B\}$. These functions are known but their parameters as $\gamma \in \Gamma$ and $\beta \in B$ are unknown. The aim is to ascertain which of the two alternatives $(f^{\gamma}(.))_{\gamma\in\Gamma}$ and $(g^{\beta}(.))_{\beta\in B}$ if any can be viewed as a family contained $h(.)$ or has a member which is a good approximate for $h(.)$. As we see, there is no trivial null hypothesis. The analysis of non-nested hypothesis is carried out with both alternatives taken in turn as the null hypothesis. There are three general approach to testing non-nested hypotheses. The Cox's test [1], [2] involves centering the log-likelihood ratio statistic under the null hypothesis. A second approach suggested by Cox [2] and improved by [3] Their basic idea is to introduce a third hypothesis in which both $(f^{\gamma}(.))_{\gamma\in\Gamma}$ and $(g^{\beta}(.))_{\beta\in B}$ are nested. A third approach, considered by [4] as encompassing procedure, focus on the ability of one model in explaining particular features of an alternative model. On the other hand Vuong's model selection test, [5], is motivated by testing that $(f^{\gamma}(.))_{\gamma\in\Gamma}$ and $(g^{\beta}(.))_{\beta\in B}$ are observationally equivalent,

using the Kullback-Leibler $\mathcal{KL}$ as a closeness measure. The focus of Vuong's test is based on a simple likelihood ratio statistics for testing the null hypothesis that the rival models are equally close to the true model against the alternative hypothesis that one model is closer in all case when the rival models are non-nested, overlapping or nested and wether both, one or neither is misspecified. The distribution free test as a model selection test introduced by Clarke [6]. The distribution-free test is that wether or not the median of the underling distribution is equal to the known value as $\mathcal{M}_0$. In section 2 we review the Vuong test and the distribution-free test in brief. Section 3 introduce the new distribution free test for model selection. Section 4 consider the asymptotic relative efficiency and the unbiasedness of the test.

2. The Vuong and the Distribution-Free tests

The Vuong and the distribution free tests are based on the $\mathcal{KL}$ divergence. Consider a sample of i.i.d. couple of variables $(Y_i, X_i), i = 1, 2, ..., n$ having joint pdf $h(.)$, $h(y, x) = h_{Y|X}(y|x)\zeta_X(x)$. Consider the model $(g^\beta(.))_{\beta\in B}$ such that $g^\beta(y, x) = g^\beta_{Y|X}(y|x)\zeta_X(x)$, the model is reduced because $\zeta_X(x)$ is assumed known. The $\mathcal{KL}$ divergence is

$$\mathcal{E}_h\{\log h_{Y|X}(Y|X)\} - \mathcal{E}_h\{\log g^\beta_{Y|X}(Y|X)\}. \tag{1}$$

For β_0 the pseudo true value of β where $\mathcal{KL}(g^\beta, h) = 0$ we have $g^\beta(.) = h(.)$, that is $\beta = \beta_0$. This risk in fact is the expectation of the log-likelihood loss of g^β relatively to $h(.)$ for observation Y condition on X, as, $\log \frac{\log h_{Y|X}(Y|X)}{\log g^\beta_{Y|X}(Y|X)}$. It is known that $\mathcal{KL}(g^\beta, h) \geq 0$. We shall say that $(g^\beta(.))_{\beta\in B}$ is closer to h than $(f^\gamma(.))_{\gamma\in\Gamma}$ if $\mathcal{KL}(g^{\beta_0}, h) < \mathcal{KL}(f^{\gamma_0}, h)$. We cannot estimate $\mathcal{KL}(g^{\beta_0}, h)$ because the entropy of h, $\mathcal{H}(h) = \mathcal{E}_h\{\log h_{Y|X}(Y|X)\}$ cannot be correctly estimated. However we can estimate the second part of (1) by $n^{-1}\{\mathcal{L}^{g^{\hat\beta_n}}_{\bar Y_n|\bar X_n}\}$ where $\hat\beta_n$ is QMLE of β_0 and $\{\mathcal{L}^{g^{\hat\beta_n}}_{\bar Y_n|\bar X_n}\}$ is the estimated log-likelihood function. However we can estimate the difference of risks as

$$\Delta(g^{\beta_0}, f^{\gamma_0}) = \mathcal{KL}(g^{\beta_0}, h) - \mathcal{KL}(f^{\gamma_0}, h) = \mathcal{E}_h\left\{\log \frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)}\right\},$$

by $n^{-1}\{\mathcal{L}^{g^{\hat\beta_n}}_{\bar Y_n|\bar X_n} - \mathcal{L}^{f^{\hat\gamma_n}}_{\bar Y_n|\bar X_n}\} = n^{-1}\{\mathcal{LR}^{g^{\hat\beta_n}, f^{\hat\gamma_n}}_n\}$, where

$$n^{-1}\{\mathcal{LR}^{g^{\hat\beta_n}, f^{\hat\gamma_n}}_n\} \xrightarrow{\mathcal{P}} \mathcal{E}_h\left\{\log \frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)}\right\},$$

see Commenges, [7].

2.1. *Vuong's model selection test*

Vuong's [5] test is constructed to test $(g^{\beta}(.))_{\beta\in B}$ against $(f^{\gamma}(.))_{\gamma\in\Gamma}$, using the $\mathcal{KL}$ divergence as a closeness measure. The focus of this approach is to test the hypothesis that the models under considerations are equally close to the true unknown model, h. The null hypothesis of Vuong's test is

$$\mathcal{H}_0 : \mathcal{E}_h \left\{ log \frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)} \right\} = 0,$$

against

$$\mathcal{H}_1 : \mathcal{E}_h \left\{ \log \frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)} \right\} > 0,$$

or

$$\mathcal{H}_1 : \mathcal{E}_h \left\{ \log \frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)} \right\} < 0.$$

Under $\mathcal{H}_0$

$$\mathcal{V}^0_n = \frac{\mathcal{LR}^{g^{\hat{\beta}_n}, f^{\hat{\gamma}_n}}_n}{\sqrt{n}\hat{\omega}_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

where

$$\hat{\omega}_n = \frac{1}{n}\sum_{i=1}^{n} \left[\log \frac{f^{\hat{\gamma}_n}_{Y|X}(Y|X)}{g^{\hat{\beta}_n}_{Y|X}(Y|X)} \right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n} \log \frac{f^{\hat{\gamma}_n}_{Y|X}(Y|X)}{g^{\hat{\beta}_n}_{Y|X}(Y|X)} \right]^2,$$

$\hat{\omega}_n$ is a consistant estimator for $\omega_* = \left\{ \mathcal{V}ar_h \left[\log \frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)} \right] \right\}^{1/2}$.

2.2. *Distribution Free-Test*

Clarke [6] consider a simple distribution-free test for non-nested model selection. This test is asymptotically more efficient than the Vuong test. His test is a modified paired sign test to the differences in the individual log-likelihood from two non-nested models. The null hypothesis of this test is

$$\mathcal{H}_0 : P \left[\log \frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)} > 0 \right] = 0.5.$$

This equation says that under the null hypothesis, the log-likelihood should be distributed around zero. The test statistic is

$$B = \sum_{i=1}^{n} 1_{(0,\infty)}(d_i)$$

where d_i is $\log \frac{f^{\hat{\gamma}_n}_{Y|X}(Y|X)}{g^{\hat{\beta}_n}_{Y|X}(Y|X)}$. It is clear that under null hypothesis, $B \overset{i.i.d.}{\sim} Bin(n, 0.5)$.

3. Linear signed rank test, modified for model selection

Since distribution free-test (DFT) introduced by Clarke [6] utilize only the sign of the difference between each likelihood ratio and the hypothesized median $\mathcal{M}_0$, the magnitudes of these observations relative to $\mathcal{M}_0$, are ignored. A test statistic which takes into account these individual relative magnitudes might be expected to give better performance. These is reasonable If we accept that the distribution of $log\frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)}$ is symmetric (especially about zero), then we can construct a new DFT, say modified DFT. Consider

$$log\frac{f^{\gamma_0}_{Y_1|X_1}(Y_1|X_1)}{g^{\beta_0}_{Y_1|X_1}(Y_1|X_1)}, ..., log\frac{f^{\gamma_0}_{Y_n|X_n}(Y_n|X_n)}{g^{\beta_0}_{Y_n|X_n}(Y_n|X_n)} \overset{i.i.d.}{\sim} F(d - \mathcal{M}).$$

Under null hypothesis $\mathcal{H}_0 : \mathcal{M} = \mathcal{M}_0$. The differences $\mathcal{D}^{lm} = log\frac{f^{\gamma_0}_{Y|X}(Y|X)}{g^{\beta_0}_{Y|X}(Y|X)} - \mathcal{M}_0$ are symmetrically distributed about zero. So that positive and negative differences of equal absolute magnitude have the same probability of occurrence, i.e.$\mathcal{P}(\mathcal{D}^{lm}_i \leq -a) = \mathcal{P}(\mathcal{D}^{lm}_i \geq a)$. We order the absolute differences $|\mathcal{D}^{lm}_1|, ..., |\mathcal{D}^{lm}_n|$ in order from smallest to largest. Their rank will be from 1 up to n, respectively. Define $\mathcal{Z}_i = 1(\mathcal{D}^{lm}_i > 0)$ for $i = 1, 2, ..., n$, where 1 stands for indicator function and

$$T_n = \sum_{i=1}^{n} \mathcal{Z}_i \text{rank}(|\mathcal{D}^{lm}_i|).$$

Now without loss of generality assume that $|\mathcal{D}^{lm}_1|, ..., |\mathcal{D}^{lm}_n|$ are order statistic, then

$$T_n = \sum_{i=1}^{n} i\mathcal{Z}_{(i)},$$

where $\mathcal{Z}_{(i)}$ is a indicator function on {The absolute value of difference with rank i is positive}. We reject $\mathcal{H}_0$ for $T_n > c$ or $T_n < d$.

4. The efficiency and the unbiasedness of test

4.1. *The efficiency*

To compare modified DFT with Vuong's test we consider the asymptotic relative efficiency (a.r.e.) of one test with respect to another test. That is given the tests T_n and W_n,

$$a.r.e.(T_n, W_n) = lim_{n\to\infty}\frac{eff(T_n)}{eff(W_n)},$$

where eff(T_n) is the efficacy of the test statistic T_n. For hypothesis $\theta = \theta_0$,

$$eff(T_n) = \frac{[\partial \mathcal{E}(T_n)/\partial\theta]^2|_{\theta=\theta_0}}{var(T_n)|_{\theta=\theta_0}}.$$

It is easy to see that

$$T_n = \binom{n}{2}^{-1} \sum\sum_{1\le i\le j\le n} 1(\mathcal{D}_i^{lm} + \mathcal{D}_j^{lm} > 0),$$

and $[\partial \mathcal{E}(T_n)/\partial\theta]|_{\theta=\theta_0} = 2\left\{\frac{f(0)}{n-1} + \int_{-\infty}^{\infty} f^2(y)\,dy\right\}$, where $y = x - \theta$, also $var(T_n)|_{\theta=\theta_0} = 3^{-1/2}$ which indicate that eff$(T_n) = 12\left\{\int_{-\infty}^{\infty} f^2(y)\,dy\right\}^2$.
It is known that the efficacy of B is $4f^2(F^{-1}(0.5))$. On the other hand the efficacy of the Vuong test could be calculate using Theorem 3.3(ii), Lemma 4.1 and Lemma 4.2. from Vuong(1989), as follows

$$\mathcal{V}_n = \frac{\sqrt{n}\left(\frac{1}{n}\mathcal{LR}_n^{g^{\hat\beta_n}, f^{\hat\gamma_n}} - \mathcal{E}_h\left\{log\frac{f_{Y|X}^{\gamma_0}(Y|X)}{g_{Y|X}^{\beta_0}(Y|X)}\right\}\right)}{\omega_*} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

and

$$\hat\omega_n \xrightarrow{a.s.} \omega_*$$

$\mathcal{V}_n$ under $\mathcal{H}_0$ is equal to $\mathcal{V}_n^0$. Using Slutsky Theorem, we have

$$\frac{\sqrt{n}\left(\frac{1}{n}\mathcal{LR}_n^{g^{\hat\beta_n}, f^{\hat\gamma_n}} - \mathcal{E}_h\left\{log\frac{f_{Y|X}^{\gamma_0}(Y|X)}{g_{Y|X}^{\beta_0}(Y|X)}\right\}\right)}{\hat\omega_n}\frac{\hat\omega_n}{\omega_*} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

but as n gets large, $\frac{\hat\omega_n}{\omega_*} \to 1$, which indicate that as $n \to \infty$

$$\mathcal{E}_h\{\mathcal{V}_n^0\} = \frac{\sqrt{n}}{\omega_*}\mathcal{E}_h\left\{log\frac{f_{Y|X}^{\gamma_0}(Y|X)}{g_{Y|X}^{\beta_0}(Y|X)}\right\},$$

and

$$\mathcal{V}ar\{\mathcal{V}_n^0\} = \frac{n\mathcal{V}ar\{\frac{1}{n}\mathcal{LR}_n^{g^{\hat\beta_n}}\}}{\omega_*^2} = 1.$$

The efficiency of the Vuong test will be $\frac{n}{\omega_*^2}$. Depending on the distribution of individual log-likelihood density the efficiency of Vuong test for uniform, $\mathcal{U}(0,1)$, normal, $\mathcal{N}(\mu,\sigma^2)$, logistic, $\mathcal{L}ogistic(\alpha,\beta)$, and double exponential, $\mathcal{D}exp(\gamma,\lambda)$, will be $12n$, $\frac{n}{\sigma^2}$, $\frac{3n}{\pi^2\beta^2}$ and $\frac{n}{2\lambda^2}$, respectively.
Table 1 shows the asymptotic relative efficiencies $(a.r.e.)$ of the three tests statistics relative to each others. The asymptotic efficiency for T_n is grater than the asymptotic efficiency of B, unless for Double exponential density which is a leptokurtic distribution with a high kurtosis. Columns 2 and 3 in Table 1 show that

in all four situations T_n is more efficient than B against the Vuong test based on $\mathcal{V}_n^0$. For Uniform density which has no tails and no peak, T_n and $\mathcal{V}_n^0$ are better than the B statistic. For the normal density, the Vuong test is better than T_n and B, but as we noted in this situation the T_n is more efficient than B statistic $(3/\pi > 2/\pi)$.

DLL	eff(B, T_n)	eff$(B, \mathcal{V}_n^0)$	eff$(T_n, \mathcal{V}_n^0)$
Uniform	$1/3 < 1$	$1/3 < 1$	1
Normal	$2/3 < 1$	$2/\pi < 1$	$3/\pi < 1$
Logistic	$3/4 < 1$	$\pi^2/12 < 1$	$\pi^2/9 > 1$
Double exp.	$4/3 > 1$	2	6

4.2. *The unbiasedness*

Consider

$$T_n = T_n(|\mathcal{D}_1^{lm}|, ..., |\mathcal{D}_n^{lm}|) = \sum_{i=1}^{n} i\mathcal{Z}_{(i)} = \sum_{i=1}^{n} i\mathcal{Z}_{(i)}^{(\mathcal{M}_0,\infty)}.$$

This statistic is used for testing $\mathcal{H}_0 : \mathcal{M} = \mathcal{M}_0$ against $\mathcal{H}_1 : \mathcal{M} > (<)\mathcal{M}_0$. If consideration is restricted to continuous distribution that are symmetric about $\mathcal{M}$ then by noting that the distribution-free test reaches its significance level, T_n provides an exact test at each of its natural level. Let $m \geq 0$

$$T_n(|\mathcal{D}_1^{lm}| + m, ..., |\mathcal{D}_n^{lm}| + m) = \sum_{i=1}^{n} i(\mathcal{Z}_{(i)}^{(\mathcal{M}_0,\infty)} + m) = \sum_{i=1}^{n} i(Z_{(i)} + m)^{(\mathcal{M}_0,\infty)},$$

Thus

$$T_n(|\mathcal{D}_1^{lm}| + m, ..., |\mathcal{D}_n^{lm}| + m) = \sum_{i=1}^{n} i\mathcal{Z}_{(i)}^{(\mathcal{M}_0-m,\infty)} \geq T_n(|\mathcal{D}_1^{lm}|, ..., |\mathcal{D}_n^{lm}|).$$

This indicates that the test has a monotone power function in $\mathcal{M}$, and therefore the test is an unbiased of $\mathcal{H}_0$ against $\mathcal{H}_1$.

References

1. D. Cox, Tests of separate families of hypothesis, in *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, (1)1961.
2. D. Cox, *Journal of the Royal Statistical Society B* **24**, 406 (1962).
3. A. Atkinson, *Ann. of Mathematical Statistics* **24**, 256 (1970).
4. N. Dastor, *Journal of Econometrica* **21**, 213 (1983).

5. Q. H. Vuong, **57**, 307 (1989).
6. K. Clarke, **24**, 1 (2007).
7. D. Commenges and A. Sayyareh, *The Annals of Applied Statistics* **3**, 1123 (2008).

Non-Nested Model Selection

A. Sayyareh (Invited)

Statistics Department, Razi University

The question of choosing a model is of course central in statistics. Model selection is estimating of different models in order to select the best one. It proceeds in two steps. The first step is to select a model between rival models and second step is to select a particular density. In this direction usually we are not in the situation without any knowledge. We have a menu of rival models which could be used to describe the data. In this paper we give an overview of statistical models, model selection and develop a methodology of model selection. The known hypothesis testing is a part of nested model selection. Our emphasize will be on the non-nested models and mis-specification. Here our focus is on asymptotic theory. We study the situation under which, the model selection procedures are asymptotically optimal for selecting a model. We will talk about some model selection criteria and some test functions suitable for model selection and compare their performances. We also give insight in the interpretation of the difference of risks in model selection.

Keywords: Akaike Information Criterion, Bayesian Information Criterion, Cox'S Test, Hypothesis Testing, Kullback-Leibler Divergence, Model Selection, Mis-Specified Models, Non-Nested Models, Vuong'S Test.

1. Introduction

An important problem in statistics concerning a sample of n independent and identically distributed observations is to test whether these observations come from a specified distribution. It means that there is a uncertainty and we have to make a decision. Decision making process under uncertainty is largely based on application of statistical data analysis for probabilistic risk assessment of our decision. In realistic situation we have only a set of data at hand and we need to build knowledge from it. The data are only crude information and not knowledge by themselves. These lead us to consider a model for data at hand. There is some essential questions in statistical inference about a mode as; What is a model? What is model selection? What are the goals of model selection? What are the methods, which methods perform better than others and in what circumstances? and so on. we begin with a known consept. A probability model is a useful concept for making sense of observations by regarding them as realizations of random variables, but the model that we can think of as having given rise to the observations is usually too complex to be described in every detail from the information available. In

literature many variants of the tests and criteria have been proposed. We may cite the Cox's test [1] suitable for model selection, Kullback-Leibler divergence [2], as a measure of discrepancy, Akaike [3] proposed the Akaike information criterion (AIC), Schwarz [4] introduced other criterion as (BIC), Vuong [5], introduced a test to model selection, Bozdogan [6] proposed complexity in model selection. In which follows we talk about the theory of model selection in brief.

2. Statistical families and statistical models

Consider $(\zeta, \mathcal{A})$ as a measurable space and $\mathcal{P}$ a subset of probabilities on it. Such a subset is called a family of probabilities. We may parameterized this family. A parameterization can be represented by a function from a set B with values in $\mathcal{P} : \beta \to P^{\beta}$. This parameterization can be denoted by $T = (P^{\beta}; \beta \in B)$. Then we have $\mathcal{P} = \{P^{\beta}; \beta \in B\}$. We call T and $\mathcal{P}$ the statistical families.
A family of probabilities on the sample space of an experiment $(\Omega, \mathcal{F})$ can be called a statistical model and a parameterization of this family will called a parameterized statistical model. If we have two parameterized statistical models $\mathcal{T} = (\mathbf{P}^{\beta}, \beta \in B)$ on $\mathcal{F}_1$ and $\mathcal{T}' = (\mathbf{P}^{\gamma}, \gamma \in \Gamma)$ on $\mathcal{F}_2$ specify the same statistical models if $\mathcal{F}_1 = \mathcal{F}_2$ and they specify the same family of probability on $(\Omega, \mathcal{F}_1)$. The pair $(Y, \mathcal{T})$ of a random variable and a parameterized statistical model induce the parameterized family of distributions on $(\mathcal{R}, \mathcal{B}) : T_Y = (P_Y^{\beta}, \beta \in B)$. Conversely, the pair (Y, T_Y) induce $\mathcal{T}$ if $\mathcal{F}_1 = \mathcal{F}$. In that case we may describe the statistical model by (Y, T_Y). Two different random variables Y and X induce two generally different parameterized families on $(\mathcal{R}, \mathcal{B})$, T_Y and T_X. For more details see [7]. Assume that there is a true, generally unknown probability $\mathbf{P}^*$. Model selection as apart of the statistical inference aims to approach $\mathbf{P}^*$.

Definition 2.1. Model $\mathcal{T}$ is well specified if $\mathbf{P}^* \in \mathcal{T}$ and is mis-specified otherwise. If it is well specified, then there is a $\beta^* \in B$ such that $\mathbf{P}^{\beta *} = \mathbf{P}^*$

3. Kullback-Leibler Risk

in decision theory, estimators are chosen as minimizing some risk function. The most important risk function is based on the Kullback-Leibler divergence [2] . Let a probability $\mathbf{P}'$ is absolutely continuous with respect to a probability $\mathbf{P}$ and $\mathcal{F}_1$ a sub-σ-field of $\mathcal{F}$ the loss using $\mathbf{P}'$ in place of $\mathbf{P}$ is the $\mathcal{L}_{\mathcal{F}}^{\mathbf{P}/\mathbf{P}'} = \log \frac{d\mathbf{P}}{d\mathbf{P}'|\mathcal{F}}$. Its expectation is

$$E_{\mathbf{P}}\{\mathcal{L}_{\mathcal{F}}^{\mathbf{P}/\mathbf{P}'}\} = KL(\mathbf{P}, \mathbf{P}'; \mathcal{F}).$$

This is the Kullback-Leibler (KL) risk. If $\mathcal{F}$ is the largest sigma-field on the space, then we omit it in the notation. If Y is random variable with p.d.f. f_Y and g_Y under $\mathbf{P}$ and $\mathbf{P}'$ respectively we have $\frac{d\mathbf{P}}{d\mathbf{P}'|\mathcal{F}} = \frac{f_Y(Y)}{g_Y(Y)}$ and the divergence of the

distribution $\mathbf{P}'$ relative to $\mathbf{P}$ can be written as

$$\mathcal{KL}(P, P') = \int \log \frac{f_Y(y)}{g_Y(y)} f_Y(y)\, d(y).$$

We have that $KL(\mathbf{P}, \mathbf{P}'; \mathcal{F}) = \mathcal{KL}(P, P')$ if $\mathcal{F}$ is the $\sigma - field$ generated by y on $(\Omega, \mathcal{F})$.

4. Model selection

Model selection is the task of choosing a model with the correct inductive bias, which in practice means selecting family of densities in an attempt to create a model of optimal complexity for the given data. Suppose a collection of data. Let $\mathcal{M}$ denote a class of these rival models. Each model $\mathcal{G} \in \mathcal{M}$ is considered as a set of probability distribution functions for the data. We note that in this framework that we do not impose that one of the candidate models $\mathcal{G}$ in $\mathcal{M}$ is a correct model. A fundamental assumption in classical hypothesis testing is that h belongs to a parametric family of densities i.e. $h \in \mathcal{G}$. To illustrate model selection, let Y be a random variable from unknown density $h(.)$. A model is assumed as possible explanation of Y, represented by $(g) = \{g(y;\beta), \beta \in B\} = (g^{\beta}(.))_{\beta \in B}$. This function is known but its parameter as $\beta \in B$ is unknown. The aim is to ascertain whethear g can be viewed as a family contained $h(.)$ or has a member which is a good approximate for $h(.)$. The $\log -$likelihood loss of g^{β} relatively to $h(.)$ for observation Y is $\log \frac{h(Y)}{g^{\beta}(Y)}$. The expectation of this loss under $h(.)$, or risk, is the $\mathcal{KL}$ divergence between g^{β} and $h(.)$ as

$$\mathcal{KL}(h, g^{\beta}) = E_h \left\{ \log \frac{h(Y)}{g^{\beta}(Y)} \right\}.$$

Let $\bar{Y} = (Y_1, Y_2, ..., Y_n)$ be identically and independently distributed random variables from unknown density $h(.)$. Two rival models are assumed as possible explanation of Y, represented by $(f^{\gamma}(.))_{\gamma \in \Gamma} = \{f(y;\gamma), \gamma \in \Gamma\}, \Gamma \subset \mathcal{R}^q$ and $(g^{\beta}(.))_{\beta \in B} = \{g(y;\beta), \beta \in B\}, B \subset \mathcal{R}^p$. These functions are known but their parameters as $\gamma \in \Gamma$ and $\beta \in B$ are unknown.The aim is to ascertain which of the two alternatives $(f^{\gamma}(.))_{\gamma \in \Gamma}$ and $(g^{\beta}(.))_{\beta \in B}$ if any can be viewed as a family contained $h(.)$ or has a member which is a good approximate for $h(.)$. As we see, there is no trivial null hypothesis.

Definition 4.1. (i) (f) and (g) are nonoverlapping if $(f) \cap (g) = \emptyset$; f is nested in (g) if $(f) \subset (g)$; (g) is well specified if there is a value $\beta_* \in B$ such that $g^{\beta_*} = h$; otherwise it is misspecified.

We assume that there is a value $\beta_0 \in B$ which minimizes $\mathcal{KL}(h, g^{\beta})$. If the model is well specified $\beta_0 = \beta_*$; if the model is misspecified, $\mathcal{KL}(h, g^{\beta}) > 0$. The Quasi Maximum Likelihood Estimator (QMLE), $\hat{\beta}_n$, is a consistent estimator of β_0, see [8,9]. The most plausible view about the statistical hypothesis is that all models are

idealization of reality, and non of them is true. But if all models are false, then the two types of errors never arises. One response to say that the null hypothesis may be approximately true, so, in which case rejecting the null hypothesis does count a mistake. Or does it? Selecting the alternative hypothesis can have more serious consequences. But we consider the alternative hypothesis to construct suitable test to model selection. It leads us to measure how far from the truth each model under null and alternative hypotheses is. This may not be possible, but we can quantify the difference of risks between two models, see [10] . The problem of testing hypothesis belonging to the same parametric family, also known as testing nested hypotheses. In classical approach, the null hypothesis is obtained as a simplified version of the alternative model. Well-known classical procedures such as those based on the likelihood ratio, Wald, and Lagrange-multiplier principal are available for testing hypotheses. When hypotheses do not belong to the same parametric family, a different approaches is necessary, since some classical procedures can not be applied.

5. Some Tests and Criteria to model selection

5.1. *Some non-nested tests*

Consider two separate families of parametric densities $\mathcal{G} \in \mathcal{M}$ and $\mathcal{F} \in \mathcal{M}$ as $\mathcal{G} = (g^{\beta}(.))_{\beta \in B}$, $\mathcal{F} = (f^{\gamma}(.))_{\gamma \in \Gamma}$ and an i.i.d. random sample from the true density $h(.)$. The model selection corresponds to the test:

$$\mathcal{H}_0^g : h(y) = g^{\beta}(y) \quad \text{for all y} \in \mathcal{Y}, \quad \beta \in B \tag{1}$$

against

$$\mathcal{H}_1^f : h(y) = f^{\gamma}(y) \quad \text{for all y} \in \mathcal{Y}, \quad \gamma \in \Gamma. \tag{2}$$

Cox's test [1] as a modified log-likelihood ratio statistic involves centering the log-likelihood ratio statistic under the null hypothesis. Cox's statistic is given by

$$\mathcal{Z}_n^{gf} = \frac{\sqrt{n}\left\{\left\{\mathcal{L}_g^{\hat{\beta}_n} - \mathcal{L}_f^{\hat{\gamma}_n}\right\} - \hat{E}_g\left\{\mathcal{L}_g^{\beta} - \mathcal{L}_f^{\gamma_g}\right\}\right\}}{\sqrt{\mathcal{V}^{gf}}}$$

where $\mathcal{L}$ stands for log likelihood function, $\mathcal{V}^{gf}$ is the asymptotic variance of the numerator of $\mathcal{Z}_n^{gf}$, $\hat{\beta}_n = argmax_{\beta \in B}\mathcal{L}_g^{\beta}$, $\hat{\gamma}_n = argmax_{\gamma \in \Gamma}\mathcal{L}_f^{\gamma}$ and $\gamma_g = argmax_{\gamma \in \Gamma}E_g\{\mathcal{L}_f^{\gamma}\}$. It is known that $\mathcal{Z}_n^{gf}$ has asymptotically standard normal distribution, see, White [9]. In Cox's framework, there is no reason to set as hypothesis as (1) and (2) since we may change the role of two hypotheses. He introduced the test statistic $\mathcal{Z}_n^{fg}$ and four rejection and acceptance regions are:
(i) : Reject both $\mathcal{H}_0^g$ and $\mathcal{H}_1^f$ if $|\mathcal{Z}_n^{gf}| \geqslant C_{\alpha}$ and $|\mathcal{Z}_n^{fg}| \geqslant C_{\alpha}$
(ii) : Reject neither $\mathcal{H}_0^g$ and $\mathcal{H}_1^f$ if $|\mathcal{Z}_n^{gf}| < C_{\alpha}$ and $|\mathcal{Z}_n^{fg}| < C_{\alpha}$
(iii) : Reject $\mathcal{H}_0^g$ but not $\mathcal{H}_1^f$ if $|\mathcal{Z}_n^{gf}| \geqslant C_{\alpha}$ and $|\mathcal{Z}_n^{fg}| < C_{\alpha}$
(iv) : Reject $\mathcal{H}_1^f$ but not $\mathcal{H}_0^g$ if $|\mathcal{Z}_n^{gf}| < C_{\alpha}$ and $|\mathcal{Z}_n^{fg}| \geqslant C_{\alpha}$

where C_α is the critical value of the standard normal at level α. Vuong's test [5] is another model selection test. In fact, Vuong's test is a relative hypothesis test where Cox's test is an absolute hypothesis test. Vuong's test uses the $\mathcal{KL}$ as a discrepancy or closeness measure to decide about $\mathcal{H}_0^g$ and $\mathcal{H}_1^f$. Vuong's test in fact is testing

$$\mathcal{H}_0^{gf} : E_h \left\{ \log \left[\frac{L_g^{\beta_\star}}{L_f^{\gamma_\star}} \right] \right\} = 0,$$

Where L stands for likelihood function. By this notation the alternative hypothesis will be one of the

$$\mathcal{H}_1^g : E_h \left\{ \log \left[\frac{L_g^{\beta_\star}}{L_f^{\gamma_\star}} \right] \right\} > 0 \quad \text{or} \quad \mathcal{H}_1^f : E_h \left\{ \log \left[\frac{L_g^{\beta_\star}}{L_f^{\gamma_\star}} \right] \right\} < 0.$$

If we accept the $\mathcal{H}_1^g$ it means that $\mathcal{G}$ is better than $\mathcal{F}$ to fit the data. In $\mathcal{KL}$ sense it means that $g^\beta(.)$ is closer than $f^\gamma(.)$ to data generating density $h(.)$. The acceptance of $\mathcal{H}_1^f$ has the same interpretation. When two rival models $g^\beta(.)$ and $f^\gamma(.)$ are non-nested. Vuong's statistic is defined as

$$\vartheta^{gf} = \frac{n^{1/2} \left\{ \frac{1}{n} \left[\mathcal{L}_g^{\hat{\beta}_n} - \mathcal{L}_f^{\hat{\gamma}_n} \right] - E_h \left\{ \log \left[\frac{L_g^{\beta_\star}}{L_f^{\gamma_\star}} \right] \right\} \right\}}{\sqrt{\omega_\star}}$$

under $\mathcal{H}_0^{gf}$ the test statistic is asymptotically distributed as a standard normal distribution and Vuong's statistic has the simpler form:

$$\vartheta^{gf} = \frac{n^{-1/2} \left\{ \mathcal{L}_g^{\hat{\beta}_n} - \mathcal{L}_f^{\hat{\gamma}_n} \right\}}{\sqrt{\hat{\omega}_n}},$$

where $\hat{\omega}_n$ is a consistent estimate of $\omega_\star = \mathcal{V}ar_h \left\{ \log \left[\frac{L_g^{\beta_\star}}{L_f^{\gamma_\star}} \right] \right\}$, as

$$\hat{\omega}_n = \frac{1}{n} \sum_{i=1}^{n} \left[\log \frac{f_{Y|X}^{\hat{\gamma}_n}(Y|X)}{g_{Y|X}^{\hat{\beta}_n}(Y|X)} \right]^2 - \left[\frac{1}{n} \sum_{i=1}^{n} \log \frac{f_{Y|X}^{\hat{\gamma}_n}(Y|X)}{g_{Y|X}^{\hat{\beta}_n}(Y|X)} \right]^2 .$$

If $\mathcal{H}_0^{gf}$ is not rejected by Vuong's test, we accept that two models are equivalent.

5.2. *Some model selection criteria*

The goal of Akaike's information criterion (AIC) is to minimize the $\mathcal{KL}$ divergence of the selected model from the true model. The Akaike's theorem provides a principaled way of trading off simplicity and fit, see [6] for details. As a essential rule of the Akaike's information criterion, the AIC rule is to select the predictive density that has the lowest estimated $\mathcal{KL}$ discrepancy, which amounts to the maximization of a penalized likelihood function. There is some difference between AIC and

classical hypothesis testing. As an important point their rationale is different. The AIC applies to nested and as well nonnested models. Akaike effectively trades off type I and type II errors in a principaled way to a tradeoff between type I and type II errors. All that is required for the comparison of models is their maximum likelihood values, the number of the parameters, and the number of data at hand. There is no need to emphasize on the null hypothesis. The AIC estimate is minus twice the expected log-likelihood and has had a huge impact on model selection, see [11]. We notice that the important part of the $\mathcal{KL}$ divergence is $E_h\{\log g^{\beta}(Y)\}$ which has an estimator as $\frac{1}{n}\sum_{i=1}^{n}\log g^{\hat{\beta}_n}(Y_i)$. It can be considered as an estimator of the divergence between the true density and the model. Akaike [3] introduced his criterion to model selection as

$$AIC = -2n\{\frac{1}{n}\sum_{i=1}^{n}\log g^{\hat{\beta}_n}(Y_i) - \hat{b}ias\} = -2\sum_{i=1}^{n}\log g^{\hat{\beta}_n}(Y_i) + 2p,$$

which is a penalized likelihood function with penalty term as the number of parameter in model. Despite of the widespread use of AIC, some believes that is too liberal and tends to select complex models (Kass and Raftery, [12]). A popular alternative model selection criterion is the Bayesian information criterion, BIC, proposed by Schwarz [4], which is defined as

$$BIC = -2\sum_{i=1}^{n}\log g^{\hat{\beta}_n}(Y_i) + p\log n,$$

Schwarz assumed that the prior probabilities of all models were equal and then derived an asymptotic expression for the likelihood of a model. Note that the BIC is similar to AIC except that it gives greater weight to simplicity by a factor of $\log n$. Where classical hypothesis testing applies, BIC is equivalent to classical test whose size slowly decreases as the sample size increases.

6. Some Recent Works

6.1. *Estimating a difference of $\mathcal{KL}$ risks*

We shall say that (g) is closer to $h(.)$ than (f) if $\mathcal{KL}(h, g^{\beta_0}) < \mathcal{KL}(h, f^{\gamma_0})$. Also, we can estimate the difference of risks $\Delta(g^{\beta_0}, f^{\gamma_0}) = \mathcal{KL}(h, g^{\beta_0}) - \mathcal{KL}(h, f^{\gamma_0})$, a quantitative measure of the difference of misspecification by $-n^{-1}(\mathcal{L}_g^{\hat{\beta}_n} - \mathcal{L}_f^{\hat{\gamma}_n})$. It is more relevant to consider the risk $E\mathcal{KL}(h, g^{\hat{\beta}_n}) = E_h\left\{\log\frac{h(Y)}{g^{\hat{\beta}_n}(Y)}\right\}$ that we call the expected $\mathcal{KL}$ risk. Thus what we really want to estimate is $\Delta(g^{\hat{\beta}_n}, f^{\hat{\gamma}_n}) = E\mathcal{KL}(h, g^{\hat{\beta}_n}) - E\mathcal{KL}(h, f^{\hat{\gamma}_n})$. Using the Akaike approximation, we obtain a simple estimator of $\Delta(g^{\hat{\beta}_n}, f^{\hat{\gamma}_n})$:

$$\hat{\Delta}(g^{\hat{\beta}_n}, f^{\hat{\gamma}_n}) = -n^{-1}\left(\mathcal{L}_g^{\hat{\beta}_n} - \mathcal{L}_f^{\hat{\gamma}_n} - (p-q)\right).$$

Thus, in contrast with AIC, $\hat{\Delta}(g^{\hat{\beta}_n}, f^{\hat{\gamma}_n})$ has an interpretation since its expectation tracks the quantity of main interest $\Delta(g^{\hat{\beta}_n}, f^{\hat{\gamma}_n})$ with good accuracy, its error is from order $o(n^{-1})$. In fact it is easy to see that

$$p_h[l_n < \Delta(g^{\hat{\beta}_n}, f^{\hat{\gamma}_n}) < u_n] = 1 - \alpha$$

where $l_n = \hat{\Delta}(g^{\hat{\beta}_n}, f^{\hat{\gamma}_n}) - z_{\alpha/2}\hat{\omega}_n$ and $u_n = \hat{\Delta}(g^{\hat{\beta}_n}, f^{\hat{\gamma}_n}) + z_{\alpha/2}\hat{\omega}_n$. We performed a simulation resembling the situation of the Depression-BMI where we have to choose between different logistic regression models. The result of simulation for $logit(h_{Y|X}(1|x_1^i, x_2^i)) = 0.5 + x_1^i + 2x_2^i$ as h, $logit(g^{\beta}_{Y|X}(1|x_1^i, x_2^i)) = \beta_0 + x_1^i + \beta_1 x_2^i$ as (g) which was wellspecified and the misspecified model (f) defiend as $logit(f^{\gamma}_{Y|X}(1|x_1^i, x_2^i)) = \gamma_0 + \sum_{j=1}^{2} \gamma_j x_{1j}^i + \gamma_3 x_2^i$ where x_{1j}^i are dummy variables indicating in which categories x_1^i fell; the categories were defined using terciles of the observed distribution of x_1^i and this was represented by two dummy variables for first and second tercile. Based on this simulation we studied the relation between BMI and depression for real data (The paquid data). In conclusion it found that there is no reason to prefer the tercile model to the linear model but there are some reasons to prefer the quadratic model to the linear model. We also studied the interaction between HIV and the immune system (analysis of the ALBI data). We fined the risks is larger than 0.28, a large difference as we have seen. This means that this difference between quiescent and activated CD4 is an important biological fact and that it must be taken into account, even though fitting the more complicated model is more challenging.
The simulation and two real problem show that the $\hat{\Delta}$ statistic and the tracking interval for the difference of risks are easy to compute and could be usefull in a wide variety of applications. For more detail see, [10].

6.2. *A comparison between some model selection criteria and tests*

Model selection is less clear when two models are equivalent. A question which arise are they closed to the unknown true model or are they far from it ? Based on a simulations, we study the results of Vuong's test, Cox's test, AIC and BIC and the ability of these four tests to discriminate between models. It remains that we decide whether or not the two rival models are two good models or two bad models which have the same distance in $\mathcal{KL}$ sense from the true model. When we say two models are equivalent, it is not clear whether two models are mis-specified or both of them are well-specified. The second case will be worst when two models overlap. Our goal is to pay attention to this problem. It seems that Cox's test is a suitable test to verify the result of Vuong's test. On the other hand the AIC and BIC are two criteria to model selection which are free of the level of test and they are not in hypothesis testing discipline. In the next part we compare these four approaches to search of the best model. To answer to our question we do a simulation study.

At the first we consider the data generating probabilities as $Lognormal(LN)$, $Gamma(\Gamma)$ and $Weibull(W)$, since they are widely used. For each case we consider two hypothesis testing problem as $H_0^g : h(.) = LN(.;\alpha_1,\alpha_2)$ against $H_1^f : h(.) = \Gamma(.;\gamma_1,\gamma_2)$ and $H_0^g : h(.) = LN(.;\alpha_1,\alpha_2)$ against $H_1^f : h(.) = W(.;\beta_1,\beta_2)$. We generate 10^4 Monte-Carlo data sets of each sample size n =50, 80, 100,200, 500. For each iteration of given sample size, we compute the test statistic for each of the criteria under related hypotheses. As expected, when the sample size increases the frequency of true decision also increases. In mis-specified case, when we use *Gamma* distribution as data generating density, the frequency of Cox result for rejecting both of hypothesis increases when the sample size increases. On the other hand we choose $Weibull$ as the data generating density with such the parameters which set this density like a *Gamma* density. As we expected Cox's test accepts the *Gamma* distribution as the selected model to describe the data. In th situation when the model is well-specified Vuong's test rejects the hypothesis contained the mis-specified one. To compare the three criteria we see that in mis-specified case Vuong's test indicates that two rival models are equivalent. As we mentioned before, when the model is well-specified result of Cox's test accepts the model as a data generating density, and the result of Vuong's test shows that this model is better than the other one. In mis-specified situations, Vuong's test indicates that two rival models are equivalent, Cox's test rejects both of two rival models. This result under Vuong's test means that two models are two bad equivalent models. Consider AIC computation for our models. In the well-specified case AIC confirms Cox and Vuong results. In the mis-specified situation AIC indicates that there is not a evidence to prefer one model as before. As we see the difference of the frequency of AIC for each of two rival models is very small. So we cannot prefer a rival model to other one, absolutely, as well as the other columns. The simulations, also, give similar results for AIC and BIC for relatively large samples. Since all models have the same number of parameters, we expected this result. The goal of AIC is to minimize $\mathcal{KL}$ divergence of alternative model from the true model. BIC is similar to AIC except that it gives greater weight to simplicity by a factor of $\log(n)$. The performance of the AIC, BIC and the likelihood function sometimes leads sometime to prefer BIC to AIC when n gets large. We find that AIC and BIC have the same behavior in non-nested model selection. An important point is that, when the hypothesis is false, Vuong's test is more sensitive to select a model. In such situations the Cox's test has the same behavior as AIC and BIC and is different from zero. The performance of both tests can also be compared for *Gamma* and $Weibull$ densities. A special nonparametric case is when we consider a symmetric distribution as a generating data model and an asymmetric distribution as a model. There is no difference between the criteria. For all sample sizes, we obtain the same conclusions. In fact the frequencies of all criteria are equal to one. But, for parametric case, we cannot guarantee that the models are really asymmetric. In our parametric example, the estimated model became asymmetric. We consider simulation for the criteria when the Weibull density is considered

as the symmetric data generating density. In a part of this case, the parametric Lognormal density is considered as the rival model, which its estimated parameters indicate that this model is asymmetric. The AIC, BIC and Cox's test have higher proportion of acceptance than Vuong's test for an asymmetric wrong model when the data generating model is symmetric. It seems Vuong's statistic has a lower probability to select a wrong model. Therefore Vuong's test is preferable to test asymmetry, see, [13].

References

1. D. Cox, *Journal of the Royal Statistical Society B* **24**, 406 (1962).
2. S. Kullback, *Information theory and statistics* (Dover, New York, 1968).
3. H. Akaike, Information theory and an extension of maximum likelihood principle, in *Second International Symposium on Information Theory, Akademia Kiado*, (Budapest, 1973).
4. G. Schwarz, *Ann. Statist.* **6**, 461 (1978).
5. Q. H. Vuong, **57**, 307 (1989).
6. H. Bozdogan, *J. Math.Psych.* **44**, 62 (2000).
7. J. Shao, *Mathematical Statistics* (Springer, USA, 2003).
8. H. White, *Econometrica* **50**, 1 (1982).
9. H. White, *Journal of Econometrics* **36**, 301 (1982).
10. D. Commenges, A. Sayyareh, L. Letenneur, J. Guedj and A. Bar-Hen, *The Annals of Applied Statistics* **3**, 1123 (2008).
11. J. DeLeeuw, Introduction to akaike (1973) information theory and an extension of the maximum likelihood principle, in *In Breakthroughs in Statistics (Kotz, S. and Johnson, N. L., eds)*, (Nagoya, Japan, 1992).
12. R. Kass and A. Raftery, *Journal of the American Statistical Association* **90**, 775 (1995).
13. A. Sayyareh, R. Obeidi and A. Bar-Hen, A simulation comparison between some model selection criteria, (2010).

Periodically Correlated Hilbertian Autoregressive Process

A. R. Soltani and M. Hashemi talkhooncheh

Department of Statistics and Operations Research, Kuwait University
Department of statistic, Shiraz university

The main of this paper is to define the periodically correlated autoregressive process in Hilbert space. We show some properties of these models such as strong law of large number together with the central limit theorem.

Keywords: Periodically Hilbertian white noise, Hilbertian periodically autoregressive process, Multivariate stationary.

1. Introduction

The Hilbertian Autoregressive model of order 1 (ARH(1)) generalizes the classical AR(1) model to random elements with values in a Hilbert space. This model was introduced by Bosq (1991), then studied by several authors such as Mourid (1993), Besse and Cardot (1996), Pumo (1999) and Mas (2003, 2007).
Periodically correlated (PC) processes in general and PC autoregressive in particular have been widely used as underlying stochastic processes for certain phenomena. The work of Hurd and miamee (1940) exhibits the importance of PC processes in applied fields. Following growing interests in operatorial statistics and practibility of functional data analysis, there is need to give attention to non-stationary Hilbertian processes. Hilbertian PC processes introduced and studied by Soltani and Shishehbor (1998, 1999). These processes assume fine time domain and spectral structurs. The processes discussed by these two authors are indeed weakly second order. Bosq in his fundamental work (2000) provides deep results on Hilbertian strongly second order autoregressive and moving average Hilbertian processes.
In this work we treate strongly second order Hilbertian PC autoregressive of order 1 (PCARH(1)), and provide results on the existence, moving average representation, strong law of large number, and central limit theorem for these processes. This article is organized as follows.
Definition of PCARH(1) process and their existence are presented in section 2. The strong law of large number together with the central limit theorem are given in section 3.

2. PCARH(1) Processes; Existence

In this article we let **H** to be a separable Hilbert space with an inner product $< \cdot, \cdot >$ and L(X) stand for the Banach space of all bounded linear operators on **H**, equipped with the operator norm. Let H^T be the cartesian product of T copies of H. If H^T is equipped with the scalar product

$$< (x_0, \cdots, x_{T-1}), (y_0, \cdots, y_{T-1}) >_T := \sum_{j=0}^{T-1} < x_j, y_j > \tag{1}$$

where $x_0, \cdots, x_{T-1}, y_0, \cdots, y_{T-1} \in \mathbf{H}$, it becomes a separable Hilbert space. The norm in H^T will be denoted by $\| \, . \, \|_T$, the space of bounded linear operators over H^T by L_T, the space of Hilbert-Schmidt (resp. nuclear) operators over H^T by S_T (resp. N_T),and the corresponding norms and scalar products by $\| \, . \, \|_{L_T}$, $\| \, . \, \|_{S_T}$, $< ., . >_{S_T}$ and $< ., . >_{N_T}$. The following definitions will be necessary.

Definition 2.1. A H-process $\{\epsilon_n, \ n \in Z\}$ is periodically H-white noise (PCHWN $(0, \sigma_t^2, T)$) if it has the following properties:
1. $E\epsilon_n = 0, \ 0 < E \parallel \epsilon_n \parallel^2 = \sigma_n^2 < \infty, \ \sigma_n^2 = \sigma_{n+T}^2$ for every n
2. $C_{\epsilon_n, \epsilon_n} = C_{\epsilon_{n+T}, \epsilon_{n+T}}$ for every n
3. $C_{\epsilon_n, \epsilon_m} = 0 \ \ for \ all \ n \neq m.$

Definition 2.2. . A sequence $\{\rho_n, \ n \in Z\}$ in L(X) is called T-periodic if $\rho_n = \rho_{n+T}, \ n \in Z$, such a sequence is completely specified by T number of bounded linear operators $\rho_0, \cdots, \rho_{T-1}$ on **H**.

Definition 2.3. Let $\{\rho_0, \cdots, \rho_{T-1}\} \subset L(X)$. We define

$$A_{0,k} = I, \ A_{i,k} = \rho_k \rho_{k-1} \cdots \rho_{k-i+1}, \ A = A_{T,T-1}. \tag{2}$$

Definition 2.4. A centered discrete time second order Hilbertian process $\mathcal{X} = \{X_n, n \in Z\}$ is called PCARH(1) with period T if it is periodically correlated and such that

$$X_n = \rho_n(X_{n-1}) + \epsilon_n \tag{3}$$

where ϵ_n is periodically H-white noise and $\{\rho_n, \ n \in Z\}$ is a T-period sequence in L(X).

There is one-to-one relationship, between PC processes and multivariate stationary processes. Indeed if $\{X_n, \ n \in Z\}$ is PC then the T-variate process $\mathbf{Y}_n = \{(X_{nT}, X_{nT+1}, \cdots, X_{nT+T-1})', \ n \in Z\}$ is stationary. Conversely if $\mathbf{Y}_n = (Y_{n,0}, \cdots, Y_{n,T-1})'$ is a T-variate stationary process then the process $X_n = Y_{m,j}, \ n = mT + j$, is PC. It is natural to ask whether the class of PCARH(1) and T-variate stationary ARH(1) coincide. We show below that the first class is indeed a proper subset of the second class.

Let $\mathcal{X} = \{X_n, n \in Z\}$ be a PCARH(1) process given by (3). Let $\mathcal{Y} = \{\mathbf{Y}_n,\ n \in Z\}$, where

$$\mathbf{Y}_n = (X_{nT}, X_{nT+1}, \cdots, X_{nT+T-1})', \tag{4}$$

and

$$\boldsymbol{\delta}_n = (\delta_{n,0}, \delta_{n,1}, \cdots, \delta_{n,T-1})',\ \ n \in Z, \tag{5}$$

where $\delta_{n,i} = \sum_{k=0}^{i} A_{k,i}\epsilon_{nT-k+i}$ for $i = 0, \cdots, T-1$ and $A_{k,i}$ for all $k = 0, 1, 2, \cdots$ are given by (2). Also for given $\rho_0, \cdots, \rho_{T-1}$, we define the following operator on $\mathbf{H}^T$

$$\boldsymbol{\Delta} = \begin{pmatrix} 0\ 0 \cdots & \rho_0 \\ 0\ 0 \cdots & \rho_1\rho_0 \\ \vdots\ \vdots\ \vdots & \vdots \\ 0\ 0 \cdots & \rho_{T-1} \cdots \rho_0 \end{pmatrix}. \tag{6}$$

We suppose that $\parallel \Delta^{j_0} \parallel_{L_T} < 1$, for some $j_0 \geq 1$ and ρ_is commute with each other. From here on this assumptions will be denoted $\mathbf{A}_1$. Let us present the following lemma, that appears to be crucial in our approach.

Lemma 2.1. *Assume stochastic processes $\mathcal{X}$ and $\mathcal{Y}$ correspond to each other through (3), (4). Then $\mathcal{X}$ is an PCARH(1) with period T, associated with $(\epsilon, \rho_0, \cdots, \rho_{T-1})$, if and only if $\mathcal{Y}$ is an $ARH^T(1)$ associated with $(\boldsymbol{\delta}, \boldsymbol{\Delta})$.*

Proof. At the first we show that $\boldsymbol{\delta}$ is a $\mathbf{H}^T$-white noise,

$$E\boldsymbol{\delta}_n = E(\delta_{n,0}, \cdots, \delta_{n,T-1})' = (E(\delta_{n,1}), \cdots, E(\delta_{n,T-1}))'$$

where $E(\delta_{n,i}) = E(\sum_{k=0}^{i} A_{k,i}\epsilon_{nT-k+i}) = \sum_{k=0}^{i} A_{k,i}E(\epsilon_{nT-k+i}) = 0$. Thus $E\boldsymbol{\delta}_n = 0,\ 0 \in \mathbf{H}^T$.
$E \parallel \boldsymbol{\delta}_n \parallel_T^2 < \infty$, because

$$E \parallel \boldsymbol{\delta}_n \parallel_T^2 = \sum_{i=0}^{T-1} E \parallel \delta_{n,i} \parallel^2 \leq 2\sum_{i=0}^{T-1}\sum_{k=0}^{i}(\max \parallel \rho_i \parallel)^k \max_i \sigma_i^2 \infty$$

because ρ_is are bounded linear operator and $E \parallel \epsilon_i \parallel^2 = \sigma_i^2 < \infty$.
It is also easy to verify that for all $n \neq m$ and $\mathbf{x}, \mathbf{y} \in \mathbf{H}^T$ we have,

$$\begin{aligned} E < \boldsymbol{\delta}_n, \mathbf{x} >< \boldsymbol{\delta}_m, \mathbf{y} > &= \sum_{i,j=0}^{T-1}\sum_{k=0}^{i}\sum_{l=0}^{j} < C_{\epsilon_{nT-k+i}, \epsilon_{mT-l+j}}(A_{k,i}^* x_i), A_{l,j}^* y_j > \\ &= 0. \end{aligned} \tag{7}$$

Therefore $\boldsymbol{\delta}_n$ and $\boldsymbol{\delta}_m$ are orthogonal, in $\mathbf{H}^T$.
Now we show that $C_{\boldsymbol{\delta}_n, \boldsymbol{\delta}_n}$ doesn't depend on n.

$$C_{\boldsymbol{\delta}_n, \boldsymbol{\delta}_n} = E < \boldsymbol{\delta}_n, . > \boldsymbol{\delta}_n = [C_{ij}]_{i,j=0,\cdots,T-1}$$

For $i \leq j$,

$$C_{ij} = \sum_{k=0}^{j} \sum_{l=0}^{i} A_{l,i} E\epsilon_{nT-k+j} \otimes \epsilon_{nT-l+i} A^*_{k,j}$$
$$= \sum_{l=0}^{i} A_{l,i} C_{\epsilon_{nT-l+i}} A^*_{l+j-i,j}$$

Also for $i > j$,

$$C_{ij} = \sum_{k=0}^{j} A_{k+i-j,i} C_{\epsilon_{nT-l+i}} A^*_{k,j}$$

Since C_{ϵ_n} is periodic on n, thus $C_{\boldsymbol{\delta}_n,\boldsymbol{\delta}_n} = C_{\boldsymbol{\delta}_{n+1},\boldsymbol{\delta}_{n+1}}$. Thus $\boldsymbol{\delta}$ is $\mathbf{H}^T$-white noise. It is clear that $\boldsymbol{\Delta}$ is linear operator from $\mathbf{H}^T$ to $\mathbf{H}^T$. We show it is bounded. Indeed using (1) we obtain that

$$\| \boldsymbol{\Delta} \|_T^2 \leq \| \rho_0 \|_L^2 + \| \rho_1\rho_0 \|_L^2 + \cdots + \| \rho_{T-1}\rho_{T-2}\cdots\rho_0 \|_L^2$$

Now we show that $\mathbf{Y}$ is a stationary process. For all $\mathbf{x}, \mathbf{y}$ in $\mathbf{H}^T$,

$$< C_{\mathbf{Y}_n,\mathbf{Y}_m}(\mathbf{x}), \mathbf{y} >_T = \sum_{j,k=0}^{T-1} < C_{X_{(n+1)T+j},X_{(m+1)T+k}}(x_j), y_k >$$
$$= E < \mathbf{Y}_{n+1}, \mathbf{x} >_T < \mathbf{Y}_{m+1}, \mathbf{y} >_T$$
$$= < C_{\mathbf{Y}_{n+1},\mathbf{Y}_{m+1}}(\mathbf{x}), \mathbf{y} >_T$$

For the only if part, we define $\epsilon_i = \delta_{n,i} - \rho_i\delta_{n,i-1}$ and $X_m = Y_{n,r}$, where $m = nT+r$. With easy computation we see that X is PCARH(1), and the proof is complete.

We now give a condition for existence of X. We will use the natural "projector" of $\mathbf{H}^T$ onto $\mathbf{H}$ defined as

$$\pi_i(x_0, \cdots, x_{T-1}) = x_i, \ (x_0, \cdots, x_{T-1}) \in \mathbf{H}^T, \ i = 0, \cdots, T-1$$

Theorem 2.1. *Under the assumption $\mathbf{A}_1$, the equation $X_n = \rho_n X_{n-1} + \epsilon_n$ has a solution given by*

$$X_{nT+i} = \sum_{j=0}^{\infty} A_{j,nT+i} \ \epsilon_{nT+i-j}, \ n \in Z$$

where $A_{j,t}$ are given by (2) and the series converges in $L^2_H(\Omega, \mathcal{A}, P)$ and with probability 1.

Proof. From property of $ARH^T(1)$ process, $\mathbf{Y}_n = \boldsymbol{\Delta}\mathbf{Y}_{n-1} + \boldsymbol{\delta}_n$ when $\| \Delta^{j_0} \|_{L_T} < 1$, for some $j_0 \geq 1$, has a unique stationary solution given by

$$\mathbf{Y}_n = \sum_{j=0}^{\infty} \boldsymbol{\Delta}^j(\boldsymbol{\delta}_{n-j}), \ n \in Z,$$

where the series converges in $L^2_{\mathbf{H}^T}(\Omega, \mathcal{A}, P)$ and with probability 1. Moreover, $(\boldsymbol{\delta}_n)$ is the innovation process of $(\mathbf{Y}_n)$. Noting that

$$X_{nT+i} = \pi_i Y_n,\ n \in Z.$$

We see that $X_{nT+i} = \pi_i \mathbf{Y}_n$

$$\begin{aligned}
&= \pi_i \sum_{j=0}^{\infty} \boldsymbol{\Delta}^j \boldsymbol{\delta_{n-j}} \\
&= \epsilon_{nT+i} + \rho_i \epsilon_{nT+i-1} + \cdots + \rho_i \cdots \rho_1 \epsilon_{nT} \\
&\quad + \sum_{j=1}^{\infty} \rho_i \cdots \rho_0 A^{j-1} [\rho_{T-1} \cdots \rho_1 \epsilon_{(n-j)T} + \cdots + \rho_{T-1} \epsilon_{(n-j)T+T-2} + \epsilon_{(n-j)T+T-1}] \\
&= \epsilon_{nT+i} + \rho_i \epsilon_{nT+i-1} + \cdots + \rho_i \cdots \rho_1 \epsilon_{nT} \\
&\quad + \sum_{k=0}^{\infty} A^k [\rho_i \cdots \rho_0 \rho_{T-1} \cdots \rho_1 \epsilon_{(n-k)T-T} + \cdots + \rho_i \cdots \rho_0 \rho_{T-1} \epsilon_{(n-k)T-2} \\
&\quad + \rho_i \cdots \rho_0 \epsilon_{(n-k)T-1}] \\
&= \epsilon_{nT+i} + \rho_i \epsilon_{nT+i-1} + \cdots + \rho_i \cdots \rho_1 \epsilon_{nT} + \sum_{k=0}^{\infty} A^k \sum_{r=i+1}^{i+T} A_{r,nT+i} \epsilon_{(n-k)T+i-r} \\
&= \sum_{j=0}^{\infty} A_{j,nT+i} \epsilon_{nT+i-j}.
\end{aligned}$$

giving the result.
It is easily follows that under the assumption (C) there are $k_0, \cdots, k_{T-1}$ such that $\sum_{i=0}^{T-1} \parallel \rho_i \parallel^{k_i} < 1$ then $\parallel \Delta^{j_0} \parallel < 1$ for some j_0.

Theorem 2.2. *Let X_n be a PCARH(1) process with period T. Suppose that there exist $\nu \in \mathbf{H}$ and $(\alpha_n) \in \mathbb{R}$ such that*

$$\rho_j^*(\nu) = \alpha_j \nu, \quad \textit{for all } j$$

and $\min_n E < \epsilon_n, \nu >^2 > 0$. Then $\{< X_n, \nu >,\ n \in Z\}$ is an PCAR(1) process that satisfies

$$< X_n, \nu > = \alpha_n < X_{n-1}, \nu > + < \epsilon_n, \nu >$$

Proof. Since X_n is PCARH(1) process, we have

$$\begin{aligned}
< X_n, \nu > = < \rho_n X_{n-1} + \epsilon_n, \nu > &= < X_{n-1}, \rho_n^*(\nu) > + < \epsilon_n, \nu > \\
&= < X_{n-1}, \alpha_n \nu > + < \epsilon_n, \nu > \\
&= \alpha_n < X_{n-1}, \nu > + < \epsilon_n, \nu >
\end{aligned}$$

It is easy to verify that $\{< \epsilon_n, \nu >,\ n \in Z\}$ is an innovation process for $< X_n, \nu >$. The scalers α_n is T-periodic. Indeed if $n = kT + r$ then

$$\rho^*_{kT+r}(\nu) = \alpha_{kT+r}\nu$$
$$\rho^*_{kT+r}(\nu) = \rho^*_r(\nu) = \alpha_r\nu$$

Thus $\alpha_{kT+r} = \alpha_r$ for any $k \in R,\ r = 0, \cdots, T-1$.
$\mathcal{X}$ is standard PCARH(1) if assumption (C) is satisfied.

3. Large sample theorem

Theorem 3.1. *Let $\mathcal{X}$ is a standard PCARH(1) and $x_0, \cdots, x_{n-1}$ be a finite sample from this model, where n is a multiple of T, n=NT. Then as $n \to \infty$,*

$$\frac{n^{\frac{1}{4}}}{(\log n)^\beta}\frac{S_n(X)}{n} \to 0,\ \ \beta > \frac{1}{2}. \tag{8}$$

Proof. We transfer $x_0, \cdots, x_{n-1}$ to a sample $\mathbf{y}_0, \cdots, \mathbf{y}_{N-1}$ from the T-variate $\mathbf{Y}$-process. According to Bosq (2000,86),

$$\frac{N^{\frac{1}{4}}}{(\log N)^\beta}\frac{S_N(\mathbf{Y})}{N} \to 0,\ \beta > \frac{1}{2}$$

But

$$\frac{N^{\frac{1}{4}}}{(\log N)^\beta}\frac{S_N(\mathbf{Y})}{N}$$
$$= (\frac{N^{\frac{1}{4}}}{(\log N)^\beta}\frac{\sum_{i=0}^{N-1} X_{iT}}{N}, \frac{N^{\frac{1}{4}}}{(\log N)^\beta}\frac{\sum_{i=0}^{N-1} X_{iT+1}}{N}, \cdots, \frac{N^{\frac{1}{4}}}{(\log N)^\beta}\frac{\sum_{i=0}^{N-1} X_{iT+T-1}}{N})'$$

Thus we obtain that

$$\frac{N^{\frac{1}{4}}}{(\log N)^\beta}\frac{\sum_{i=0}^{NT-1} X_i}{N} = \frac{N^{\frac{1}{4}}}{(\log N)^\beta}\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{T-1} X_{iT+j}}{N} \to 0,\ \ \beta > \frac{1}{2}.$$

On the other hand,

$$\frac{(NT)^{\frac{1}{4}}}{(\log NT)^\beta} \mid \frac{\sum_{i=0}^{NT-1} X_i}{NT} \mid \leq \frac{(NT)^{\frac{1}{4}}}{(\log N)^\beta} \mid \frac{\sum_{i=0}^{NT-1} X_i}{N} \mid$$

Thus we arrive at (8). The proof is complete.

Lemma 3.1. *$(I_T - \mathbf{\Delta})$ is invertiable if and only if $(I - A_{T,T-1})$ is invertiable in* **H**.

Proof. Let us consider the case T=2. We have

$$I_2 - \mathbf{\Delta} = \begin{pmatrix} I & -\rho_0 \\ 0 & I - \rho_1\rho_0 \end{pmatrix}$$

and invertibility of $I_2 - \mathbf{\Delta}$ means exists matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ such that

$$(I_2 - \mathbf{\Delta}) \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} (I_2 - \mathbf{\Delta}) = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

Where α, β, γ *and* δ are in L.
With easy calculation we can show that

$$(I_2 - \mathbf{\Delta})^{-1} = \begin{pmatrix} I & \rho_0(I - \rho_1\rho_0)^{-1} \\ 0 & (I - \rho_1\rho_0)^{-1} \end{pmatrix}$$

It means that $(I_2 - \mathbf{\Delta})^{-1}$ exists if $I - \rho_1\rho_0$ is invertiable.
Conversely, if $I - \rho_1\rho_0$ is invertiable, one may set

$$\alpha = I, \;\; \beta = \rho_0(I - \rho_1\rho_0)^{-1}, \;\; \gamma = 0, \;\; \delta = (I - \rho_1\rho_0)^{-1}$$

Then it is easy to verify that the matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ is inverse of $I_2 - \mathbf{\Delta}$.
If $T > 2$, a recursive argument leads to the general statement.
We now state Central Limit Theorem.

Theorem 3.2. *Let X be a standard PCARH(1) associated with periodically independent and identical distributed white noise and such that $(I_T - A_{T,T-1})$ is invertiable Then*

$$\frac{S_n}{n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma)$$

Proof. By the theorem 3.10 of Bosq we obtain

$$S_{\mathbf{Y}} = \frac{\mathbf{Y}_0 + \mathbf{Y}_1 + \cdots + \mathbf{Y}_{N-1}}{\sqrt{N}} \xrightarrow{\mathcal{D}} \mathcal{N}'(0, \mathbf{\Gamma})$$

where

$$\mathbf{\Gamma} = (I_T - \mathbf{\Delta})^{-1} C_{\boldsymbol{\delta}_n} (I_T - \mathbf{\Delta}^*)^{-1}$$

Nothing that

$$S_{NT} = \sum_{i=0}^{NT-1} X_i = A' S_{\mathbf{Y}}$$

where $A = (I, I, \cdots, I)'$. We get

$$\frac{S_{NT}(X)}{\sqrt{NT}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{A'\mathbf{\Gamma}A}{T})$$

References

1. A. Mas, Weak convergence in the functional autoregressive model, *J. Mult. Anal.*, **98**, 1231-1261 (2007).
2. A. R. Soltani, Z. Shishehbor, Weakly periodic sequences of bounded linear transformations: A spectral characterization, *J. Georgian Math.*, **6**, 91-98 (1999).
3. A. R. Soltani, Z. Shishehbor, A spectral representation for weakly periodic sequences of bounded linear transformations, *Acta. Math. Hungar*, **80**, 265-270 (1998).
4. B. Pumo, Prediction of continuous time processes by C[0,1]-valued autoregressive process, *Stat. Infer. Stoch. Processes*, **3**, 1-13 (1999).
5. D. Bosq, it Linear Processes in Function Spaces. Theory and Applications, Lecture Notes in Statistics, vol. 149,(Springer, Berlin, 2000).
6. H. L. Hurd, A. Miamee, *Periodically Correlated Random Sequences. Spectral Theory and Practice*, (Wiley).
7. P. Besse, H. Cardot, Approximation spline de la prevision d'un processus fonctionnel autoregressif d'ordre, *Canad. J. Statist.*, **24**, 467-487 (1996).
8. T. Murid, Processus autoregressifs d'ordre superieur, *Acad. Sci.*, 1167-1172 (1993).
9. W. Rudin, *Functional analysis*, (New York, McGraw Hill, 1976).

On The Generalized Cauchy And Truncated Generalized Cauchy Distributions

A. R. Soltani, A. R. Nematollahi and L. Tafakori

Department of Statistics, Shiraz university

The generalized Cauchy distribution is introduced by Rider (1957). He studied the properties of samples from these kind of distributions only in a special case, say $m = 1$. In this paper, we have studied the properties of generalized Cauchy distribution for $m \geq 1$. A new representation for the characteristic function of a generalized Cauchy distribution is derived by the well-known generalized non-symmetric Linnik distributions. The truncated generalized Cauchy distribution or cutoff generalized Cauchy distribution for $m \geq 1$, is also introduced. Particularly, the representation of the characteristic function for these kind of distributions is derived.

Keywords: Generalized Linnik Distribution, Generalized Cauchy Distribution, Dual Probability Density.

1. Introduction

Recently, the modeling and predicting of the financial data has attracted attention of numerous researchers. Stable distributions provide approximations for sums of i.i.d. random variables that have heavy tails, and thus seemed appropriate for modeling financial data. A stable distribution with parameter α, is often called $\alpha-$stable and is reduced to Cauchy distribution if $\alpha = 1$. The Cauchy distribution is of considerable interest for several reasons. It is the distribution of the quotient of two independent normal random variables. The distribution of means of samples from it, is exactly the same as the parent distribution itself. Moreover, the median of a sample from a Cauchy distribution is more efficient than the mean in estimating the center of the population. Rider (1957) introduced the generalized Cauchy distribution, but he studied the properties of samples from these kind of distributions only in a special case, say $m = 1$. In this paper, we have studied the properties of generalized Cauchy distribution for $m \geq 1$. The truncated generalized Cauchy distribution or cutoff generalized Cauchy distribution for $m \geq 1$, is also introduced. Particularly, the representation of the characteristic function for these kind of distributions is derived.

2. The generalized Cauchy distribution

The probability density function of a random variable X with generalized Cauchy (GC) distribution is of the form

$$f(x) = \frac{\Gamma(m)}{b\Gamma(\frac{1}{2})\Gamma(m-\frac{1}{2})}(1 + \left(\frac{x- a}{b}\right)^2)^{-m}, m \geq 1, x \in R, \tag{2.1}$$

where $a \in R$ and $b > 0$, (See Rider (1957) and Evans and Hastings (2000)). For $m = 1$, this distribution corresponds to the Cauchy distribution and for $a = 0$, it reduces to a t-student distribution with $(2m - 1)$ degrees of freedom , multiplied by $b(2m - 1)$.

By taking $a = -\cos(\rho\pi)$ and $b = \sin(\rho\pi)$, the following representation of (2.1) will follow

$$f(x) = \frac{\sin^{2m-1}(\rho\pi)\,\Gamma(m)}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})}(\sin^2(\rho\pi) + (x + \cos(\rho\pi))^2)^{-m} , m \geq 1,\ 0 < \rho < 1. \tag{2.2}$$

Since the distribution is symmetric about $-\cos(\rho\pi)$, its odd moments, when they exist, are zero. The mean, median and mode are also equal to $\mu = -\cos(\rho\pi)$.

The r-th moment (for even r) about the mean is

$$E(X - \mu)^r = \frac{\Gamma(\frac{r+1}{2})\,\Gamma(m - \frac{r+1}{2})}{\Gamma(\frac{1}{r})\,\Gamma(m - \frac{1}{r})}, r < 2m - 1.$$

In the following theorem , we present a closed form of the corresponding characteristic function $\phi_X(t)$ of the density given by (2.2).

Theorem 2.1 For $0 < \rho < 1$, the characteristic function of the generalized Cauchy distribution given by (2.2), can be represented by

$$\phi_X(t) = \begin{cases} e^{-it\cos(\rho\pi)}e^{-\sin(\rho\pi)|t|}, & m = 1 \\ \frac{\Gamma(m)e^{-it\cos(\rho\pi)}}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})}(2\pi i)\frac{\Psi^{(m-1)}(i)}{(m-1)!}, & m \geq 2 \end{cases},$$

where $\Psi(z) = \frac{e^{it(\sin(\rho\pi))z}}{(z+i)^m}$, and $\Psi^{(m-1)}$ denotes the $(m-1)$th derivative of Ψ.

Proof. In the complex x-plane , consider the region

$$Q_R = \{x = u + iv \ : |x| < R\,, v > 0\ \}\,,$$

and define $x = |x|\,e^{i\phi}, 0 \leq \phi = \arg x \leq \pi$.

The integrand is analytic in the closure of Q_R except the simple poles at $x = \pm i$ .By using Cauchy's residue theorem we have

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx}\, f_X(x)dx = \frac{\sin^{2m-1}(\rho\pi)\,\Gamma(m)}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})} \int_{-\infty}^{\infty} \frac{e^{itx}}{(x^2+2x\cos\rho\pi+1)^m}$$

$$= \frac{\sin^{2m-1}(\rho\pi)\,\Gamma(m)}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})} \int_{-\infty}^{\infty} \frac{e^{itx}}{((x+\cos\rho\pi)^2+\sin^2\rho\pi)^m}dx$$

$$= \frac{\sin^{2m-1}(\rho\pi)\,\Gamma(m)}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})} \frac{e^{-it\cos(\rho\pi))}}{\sin^{2m-1}(\rho\pi)} \int_{-\infty}^{\infty} \frac{e^{it(\sin(\rho\pi)u}}{(u^2+1)^m}du$$

$$= \frac{\Gamma(m)e^{-it\cos(\rho\pi))}}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})} \int_{-\infty}^{\infty} \frac{\frac{e^{it(\sin(\rho\pi)u}}{(u+i)^m}}{(u-i)^m}du$$

$$= \begin{cases} e^{-it\cos(\rho\pi)}(2\pi i)\Psi(i) & m=1 \\ \frac{\Gamma(m)e^{-it\cos(\rho\pi)}}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})}(2\pi i)\frac{\Psi^{(m-1)}(i)}{(m-1)!} & m\geq 2 \end{cases}$$

$$= \begin{cases} e^{-it\cos(\rho\pi)}e^{-\sin(\rho\pi)|t|} & m=1 \\ \frac{\Gamma(m)e^{-it\cos(\rho\pi)}}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})}(2\pi i)\frac{\Psi^{(m-1)}(i)}{(m-1)!} & m\geq 2 \end{cases}$$

The proof is now completed .□

In the next section, a new representation for the characteristic function of a generalized Cauchy distribution is derived by the well-known generalized non-symmetric Linnik distributions (GNSL) (see Kozubowski [2000] for more details about GNSL).

3. A representation for the characteristic function of the GC distribution by using GNSL distributions

It is a challenging task to derive the characteristic function of the GC distributions by using generalized non-symmetric Linnik distributions (GNSL), which is also called .the geo-stable distribution in the literature. Kozubowski [2000] provide a representation of the GNSL density in terms of the corresponding $\alpha-$stable density. In the next theorem, we first present a direct integral representation for the density of GNSL.

Theorem 3.1. For $|\theta| < 1$, the density of $GNSL(m,\theta)$, $f(x)$, can be represented by

$$f(x) = \frac{1}{\pi}\int_0^{\infty} \frac{e^{-y\,|x|}\,\sum_{k=0}^{m}\binom{m}{k}y^k\sin k\pi\frac{1}{2}(1+\theta\ sign(x))}{\left[y^2+2y\cos\frac{\pi}{2}(1+\theta\ sign(x))+1\right]^m}\,dy, x\neq 0,$$

where $sign(x)$ is equal to 1 for $x > 0$, and is equal -1 for $x < 0$.

Proof. First let $x < 0$, and split $f(x)$ into two integrals as follows

$$I_1(x) := \int_0^\infty e^{-itx} \phi_X(t)\ dt, \tag{3.1}$$

and $I_1(x) + I_2(x) = 2\pi f(x)$.

In the complex t-plane, consider the region

$$Q_R = \{t = u + iv\ : |t| \le R\,, u \ge 0\,, v \ge 0\,\},\ R > 1,$$

and define $t = |t|\ e^{i\phi}, 0 < \phi = \arg t < \frac{\pi}{2}$.

The function $\frac{e^{-itx}}{\left[1+te^{-i\frac{\pi}{2}\ \theta}\right]^m}$ is continuous in Q_R and analytic in the interior of Q_R. Let $q(R)$ be the boundary of Q_R and $C(R) := \{t = u + iv\ : |t| = R\,, u \ge 0\,, v \ge 0\,\}$. Then by Cauchy theorem, we have

$$\int_{q(R)} \frac{e^{-itx}}{\left[1 + te^{-i\frac{\pi}{2}\ \theta}\right]^m} dt = 0.$$

Therefore,

$$\int_0^R \frac{e^{-itx}}{\left[1 + te^{-i\frac{\pi}{2}\ \theta}\right]^m} dt + \int_{C(R)} \frac{e^{-itx}}{\left[1 + te^{-i\frac{\pi}{2}\ \theta}\right]^m} dt - i\int_0^R \frac{e^{xy}}{\left[1 + ye^{i\frac{\pi}{2}\ (1-\theta)}\right]^m} dy = 0.$$

The integral over $C(R)$ tends to zero as R tends to infinity. Hence,

$$I_1(x) = i\int_0^\infty \frac{e^{xy}}{\left[1 + ye^{i\frac{\pi}{2}\ (1-\theta)}\right]^m} dy. \tag{3.2}$$

Then $I_1(x)$ is an exponentially convergent integral, since $x < 0$.

To obtain an analogous representation for

$$I_2(x) = -i\int_0^R \frac{e^{xy}}{\left[1 + ye^{i\frac{\pi}{2}\ (\theta-1)}\right]^m} dy.$$

Consider now region

$$P_R = \{t = u + iv\ : |t| \le R\,, u \ge 0\,, v < 0\,\},\ R > 1.$$

In the complex t-plane and define $t = |t|\, e^{i\phi - 2\pi i},\ \frac{3\pi}{2} < \phi = \arg t < 2\pi$.

Applying to function $I_2(x)$ and set P_R , the arguments similar to those used in above for $I_1(x)$, leads to the representation

$$I_2(x) = -i \int_0^R \frac{e^{xy}}{\left[1 + ye^{i\frac{\pi}{2}\,(\theta-1)}\right]^m} dy.$$

Combining the representations for $I_1(x)$ and $I_2(x)$, we obtain

$$f(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{yx} \sum_{k=0}^{m} \binom{m}{k} y^k \sin k\pi \frac{1}{2}(1-\theta\,)}{\left[y^2 \;+\; 2y\cos\frac{\pi}{2}(1-\theta\,)+1\right]^m}\, dy,\; x<0. \tag{3.3}$$

Proceeding in a similar way for $x > 0$, we obtain

$$f(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{-yx} \sum_{k=0}^{m} \binom{m}{k} y^k \sin k\pi \frac{1}{2}(1+\theta\,)}{\left[y^2 \;+\; 2y\cos\frac{\pi}{2}(1+\theta\,)+1\right]^m}\, dy,\; x>0. \tag{3.4}$$

The proof is now completed .□

Rosseberg (1990) introduced the probability densities which are proportional to the characteristic functions and therefore are positive definite. Positive definite densities $f(x)$ are continuous , symmetric, and satisfy $0 \leq f(x) < f(0),\ x \neq 0$. Their most attractive property, however, is the existence of an adjoint, reciprocal or dual probability density $\widehat{f}$. Denoting the corresponding characteristic functions by ϕ and $\widehat{\phi}$, the densities f and $\widehat{f}$ are connected by the fundamental relations

$$\phi(t) = \frac{\widehat{f}(t)}{\widehat{f}(0)}\ ,\ \widehat{\phi}(t) = \frac{f(t)}{f(0)},\ 2\pi f(0)\widehat{f}(0) = 1,$$

see Rosseberg (1990) for details.

Now, we are in position to provide an alternative transparent derivation of the characteristic function of the GC distributions, by using the density of GNSL. Let $a = -\cos(\frac{\theta\pi}{2})$ and $b = \sin(\frac{\theta\pi}{2}), 0 < \theta < 1$ in (2.1) and let $X \sim GNSL(m,\theta)$. Consider also a random variable$Y$ independent of X such that $-Y \overset{d}{=} X$. We are interested to the distribution of $Z = X + Y$. Evidently, the characteristic function of Z will be

$$\phi_Z\,(t) = \frac{1}{\left(1+|t|\,e^{i\frac{\pi}{2}\;\theta\;sign(t)}\right)^m} \times \frac{1}{\left(1+|t|\,e^{-i\frac{\pi}{2}\;\theta\;sign(t)}\right)^m} = \frac{1}{\left|1+|t|\,e^{-i\frac{\pi}{2}\;\theta\;sign(t)}\right|^{2m}},\ t \in R,$$

and the density function of Z is

$$f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\, dx$$

$$= \int_0^{+\infty} [\frac{1}{\pi} \int_0^{\infty} \frac{2}{(\eta+\nu)} \frac{e^{\nu\, z} \sum_{k=0}^{m} \binom{m}{k} \nu^k \sin k\pi \frac{1}{2}(1+\theta)}{\left|1+\nu e^{i\frac{\pi}{2}(1+\theta)}\right|^{2m}}\, d\nu$$

$$\times \frac{1}{\pi} \int_0^{\infty} \frac{e^{\eta\, z} \sum_{k=0}^{m} \binom{m}{k} \eta^k \sin k\pi \frac{1}{2}(1-\theta\ )}{\left|1+\eta e^{i\frac{\pi}{2}(1-\theta)}\right|^{2m}}\, d\eta],$$

where $z \in R$.

Since

$$f_Z(0) = \int_0^{+\infty} [\frac{1}{\pi} \int_0^{\infty} \frac{2}{(\eta+\nu)} \frac{\sum_{k=0}^{m} \binom{m}{k} \nu^k \sin k\pi \frac{1}{2}(1+\theta\ )}{\left|1+\nu e^{i\frac{\pi}{2}(1+\theta)}\right|^{2m}}\, d\nu$$

$$\times \frac{1}{\pi} \int_0^{\infty} \frac{\sum_{k=0}^{m} \binom{m}{k} \eta^k \sin k\pi \frac{1}{2}(1-\theta\ )}{\left|1+\eta e^{i\frac{\pi}{2}(1-\theta)}\right|^{2m}}\, d\eta]\ ,$$

we arrive at the following representation for the characteristic function of the GC random variables

$$\phi_{GC}\,(z) = \frac{f_Z(z)}{f_Z(0)}.$$

4. The truncated generalized Cauchy distribution

The random variable X is said to be the truncated generalized Cauchy (TGC) distribution, if its density function has the following form,

$$f_X(x) = \tfrac{1}{A} \tfrac{1}{(x^2+2x\cos\rho\pi+1)^m}\ ,\ m>0,\ 0<\rho<1, \tag{4.1}$$

where

$$A = \left\{ \frac{-1}{2m-1} \sum_{k=0}^{m-2} \frac{(2m-1)\,(2m-3)\,...\,(2m-1-2k)}{(m-1)\,(m-2)\,...\,(m-1-k)} \frac{\cos\rho\pi}{2^{k+2}\sin^{2k+1}\rho\pi} \right\}$$

$$+ \left\{ \frac{(2m-3)\times(2m-5)\times ...5\times 3\times 1}{(m-1)\times(m-2)\times ...2\times 1} \frac{(\rho\pi)}{2^{m-1}\sin^{2m-1}\rho\pi} \right\}.$$

The moments of TGC distribution is

$$E(X^{r-1}) = \int_0^\infty \frac{x^{r-1}}{A\,(1+2x\,\cos\rho\pi\,+x^2)^m}dx$$
$$= \frac{1}{A}2^{m-\frac{1}{2}}(\sin\rho\pi)^{\frac{1}{2}-m}t\;\Gamma(m+\frac{1}{2})\;B(r\;,2m-r)\times P^{\frac{1}{2}-m}_{r-m-\frac{1}{2}}(\cos\rho\pi)$$

where $0<\rho\pi<\pi\;, 0<Re(r)<Re(2m)$, and

$$P^s_m(\cos\phi)=\sqrt{\frac{2}{\pi}}\frac{\sin^s\phi}{\Gamma(\frac{1}{2}-s)}\int_0^\phi\frac{\cos(\upsilon+\frac{1}{2})\;t}{(\cos t-\cos\phi)^{s+\frac{1}{2}}}dt,$$

where $0<\phi<\pi, Re(s)<\frac{1}{2}$.

Theorem 4.1 For $0<\theta<1$, the characteristic function of the truncated generalized Cauchy distribution can be represented by

$$\phi_X(t)\;=\frac{i}{A}\int_0^\infty\frac{e^{-\upsilon|t|}\;sgn(t)}{(1+\upsilon\;e^{i\frac{\pi}{2}(\theta+1)sgn(t)})^m(1+\upsilon\;e^{-i\frac{\pi}{2}(\theta-1)sgn(t)})^m}d\nu.$$

Proof. We have

$$\phi_X(t)=\int_0^\infty e^{itx}\;f_X(x)dx$$
$$=\int_0^\infty\frac{1}{A}\frac{e^{itx}}{(x^2+2x\cos\frac{\theta\pi}{2}+1)^m}dx$$
$$=\frac{1}{A}\int_0^\infty\frac{\frac{e^{itx}}{(1+x\;e^{i\frac{\pi}{2}\theta})^m}}{(1+x\;e^{-i\frac{\pi}{2}\theta})^m}dx.$$

Then in the complex x-plane, we consider the region

$$Q_R=\{x=u+iv\;:|x|\le R\;,u\ge 0\;,v\ge 0\;\}.$$

and define $x=|x|\;e^{i\phi}, 0<\phi=\arg x<\frac{\pi}{2}$.

According to the Cauchy theorem , we have

$$\int_{C(R)}\frac{\frac{e^{itx}}{(1+x\;e^{i\frac{\pi}{2}\theta})^m}}{(1+x\;e^{-i\frac{\pi}{2}\theta})^m}dx=0$$

Therefore

$$\frac{1}{A}\int_0^R \frac{\frac{e^{itx}}{(1+x\ e^{i\frac{\pi}{2}\theta})^m}}{(1+x\ e^{-i\frac{\pi}{2}\theta})^m}dx + \frac{1}{A}\int_{C(R)} \frac{\frac{e^{itx}}{(1+x\ e^{i\frac{\pi}{2}\theta})^m}}{(1+x\ e^{-i\frac{\pi}{2}\theta})^m}dx - \frac{i}{A}\int_0^R \frac{\frac{e^{-\upsilon t}}{(1+\upsilon\ e^{i\frac{\pi}{2}(\theta+1)})^m}}{(1+\upsilon\ e^{-i\frac{\pi}{2}(\theta-1)})^m}d\nu = 0.$$

The integral over $C(R)$ tends to zero as R tends to infinity. Hence,

$$I_1(x) = \frac{i}{A}\int_0^\infty \frac{\frac{e^{-\upsilon t}}{(1+\upsilon\ e^{i\frac{\pi}{2}(\theta+1)})^m}}{(1+\upsilon\ e^{-i\frac{\pi}{2}(\theta-1)})^m}d\nu, t>0$$

Consider now region

$$P_R = \{x = u+iv : |x| \le R, u \ge 0, v \le 0\}.$$

In the complex x-plane and define $x = |x|\ e^{i\phi-2\pi i},\ \frac{3\pi}{2} < \phi = \arg x < 2\pi$.

The similar arguments to those used in above for $I_1(x)$ leads to the following representation

$$I_2(x) = \frac{-i}{A}\int_0^\infty \frac{\frac{e^{\upsilon t}}{(1+\upsilon\ e^{i\frac{\pi}{2}(\theta-1)})^m}}{(1+\upsilon\ e^{-i\frac{\pi}{2}(\theta+1)})^m}d\nu, t<0.$$

Combining the representations for $I_1(x)$ and $I_2(x)$, we obtain

$$\begin{aligned}\phi_X(t) &= \int_0^\infty e^{itx}\ f_X(x)dx = \int_0^\infty \frac{1}{A}\frac{e^{itx}}{(x^2+2x\cos\frac{\theta\pi}{2}+1)^m}dx \\ &= \frac{1}{A}\int_0^\infty \frac{\frac{e^{itx}}{(1+x\ e^{i\frac{\pi}{2}\theta})^m}}{(1+x\ e^{-i\frac{\pi}{2}\theta})^m}dx \\ &= \frac{i}{A}\int_0^\infty \frac{e^{-\upsilon|t|}\ sign(t)}{(1+\upsilon\ e^{i\frac{\pi}{2}(\theta+1)sign(t)})^m(1+\upsilon\ e^{-i\frac{\pi}{2}(\theta-1)sign(t)})^m}d\nu.\end{aligned}$$

The proof is now completed .□

References

1. M.Evans, and N.Hastings,and B.Peacock, Statistical distributions(2000).
2. T.J. Kozubowski, Exponential mixture representation of geometric stable distributions, *Ann. Inst. Statist. Math.*, **52**, 2, 231-238 (2000).
3. H. J. Rosseberg, Positive definite probability densities. *Theory of probability and its applications*, **35**, 169-174 (1990).

Criteria For Model Discrimination And Parameter Estimation Performance Of Search Designs

H. Talebi (Invited) and N. Esmailzadeh

Department of Statistics, University of Isfahan

Department of Statistics, University of Kurdistan

Since introduction of search designs by Srivastava (1975), these designs have been constructed by many researchers. There also have been developed some criteria in order to evaluate the search ability of these designs. These criteria are reviewed in this paper. Search designs, which are efficient with respect to search, may be poor in parameter estimation. In this paper, we propose a dual-task criterion to deal with both search and estimation performance of the search designs. The criterion is implied to some designs and the results are given.

Keywords: Search Designs, Search Linear Model, Search Probability , Model Discrimination

1. Introduction

In a 2^m factorial experiment, let $\beta' = (\beta_1', \beta_2')$ be a $p \times 1$ vector of factorial effects. $\beta_1(\nu_1 \times 1)$ is vector of fixed unknown parameters and $\beta_2(\nu_2 \times 1)$ is partially known. It is known that at most k elements of β_2 are nonzero but those elements are unknown. The problem is search for nonzero elements of β_2 to identify and estimate them along with estimating the effects in β_1. This is so-called search problem introduced by Srivastava (1975), though deals with both search and estimation. The following search linear model is considered to solve the problem,

$$y = X_1\beta_1 + X_2\beta_2 + e, \quad Var(e) = \sigma^2 I, \tag{1}$$

where $y(N \times 1)$ is a vector of observations, $X_i(N \times \nu_i)$ are known design matrices, $e(N \times 1)$ is an error vector, σ^2 is the error variance and I_N is the identity matrix of order N. A design which is able to solve this problem is so-called a search design. For every $N \times 2k$ submatrix X_{22} of X_2, the search design must satisfy the following condition (Srivastava, 1975),

$$\text{rank}(X_1; X_{22}) = \nu_1 + 2k. \tag{2}$$

For the noiseless case, $\sigma^2 = 0$, this is necessary and sufficient condition. However, for the noisy case, $\sigma^2 > 0$, it is not sufficient but is still necessary. Several

researchers have been studied and obtained the search designs for the noiseless case. Some of those are Ghosh (1981), Onishi and Shirakura (1985), Gupta (1984, 1988), Shirakura (1991) and Ghosh and Talebi (1993). One can see Ghosh(1996) for more details on design constructions. For noisy case, Srivastava (1975) suggested to calculate sum of squared error (SSE) for all of possible $\binom{\nu_2}{k}$ candidate models, include β_1 in common and different $\beta_{2i}(k \times 1)$ in β_2 from (1). That is

$$M_i : y = X_1\beta_1 + X_{2i}\beta_{2i} + e, \quad i = 1, 2, ..., \binom{\nu_2}{k} \tag{3}$$

where $\beta_{2i} \in \beta_2$ and X_{2i} are columns in X_2 corresponding to β_{2i}. The model with minimum SSE is chosen over the alternatives as the true model, M_0, including $\beta_{20} \in \beta_2$. However, one may come up with a wrong decision , due to stochastic property of SSE. Shirakura et al. (1996) considered the search ability of a design in discriminating the true model. They defined the search probability (SP) to measure such an ability. That is, for SSE_0 and SSE_i of true and any alternative models, respectively, the SP is $P(SSE_0 < SSE_i|\beta_{20}, \beta_{2i}, \sigma^2)$. They also obtained an exact expression of SP for $k = 1$, when errors are normally distributed. Namely, for $N \times N$ matrix $Q = X_1(X_1'X_1)^{-1}X_1$ and $r_{2i} = x_{2i}'[I - Q]x_{2i}$, in which x_{2i} is $N \times 1$ column of X_2 corresponding to $\beta_{2i} \in \beta_2$, the SP is reduced to

$$G(\beta_{20}, \beta_{2i}, \rho) = 1 - \Phi(c_1\rho) - \Phi(c_2\rho) + 2\Phi(c_1\rho)\Phi(c_2\rho), \tag{4}$$

where $\Phi(\cdot)$ is the standard normal cdf, $c_1 = \sqrt{\frac{r_{20}}{2}(1 - x)}$, $c_2 = \sqrt{\frac{r_{20}}{2}(1 + x)}$, $x = x_{20}'[I - Q]x_{2i}/[r_{20}r_{2i}]^{1/2}$ and $\rho = \beta_{20}/\sigma$. Clearly, G depends on the design through c_1 and c_2 and is a symmetric function of unknown effect size ρ. By symmetric property of G in ρ, we take ρ as a non-negative constant for case $k = 1$, from now on. For such a ρ, given r_{20} and $|x| < 1$, function G is an increasing function of c_1 (Ghosh and Teschmacher, 2002). Note that for each $\beta_{20} \in \beta_2$ there are $\binom{\nu_2}{k} - 1$ alternative sets $\beta_{2i} \in \beta_2$, " of which at least one effect in β_{2i} is not of β_{20}. We denote this by $\mathcal{A}(\beta_{20}, \beta_2)$. Consider the $\nu_2 \times \nu_2$ matrix with elements $G(\beta_{20}, \beta_{2i}, \rho)$, whose columns are corresponding to all $\beta_{20} \in \beta_2$ and rows to the alternatives $\beta_{2i} \in \beta_2$. This is so-called search probability matrix (SPM). Clearly, the diagonal elements of SPM are not of interest. Based on (4) and by considering the SPM, several authors have been developed some criteria for measuring and comparing the search performance of designs. They are Shirakura et al. (1996), Ghosh and Teschmacher (2002) and Talebi and Esmailzadeh (2009). The results of these researchers are briefly reviewed and presented in the following section. For $k \geq 1$, recently, Talebi and Esmailzadeh (2010) proposed two criteria based on Kullback - Leibler distance to do the task. Note that, all of above mentioned criteria measure the search performance of designs, but not of considering the efficiency of the designs in parameter estimation for underlying models. So, one may come up with efficient search design which is poor in estimation properties. Construction of a design with a dual efficiency, i.e. model discrimination and parameter estimation, is back to Hill et al. (1968). Later on, this problem has been tackled by Dette

(1993), Biswass and Chaudhuri (2002) and Waterhouse et al. (2004). The most recent works have been done by Atkinson (2008) and Tommasi (2009), in which they presented new criteria for obtaining optimum design, which is also efficient in model discrimination. In Section 3 we will develop new criteria to do the dual task in the context of search linear model (1). This allows us to balance between these two aspects of performance efficiency of designs.

2. A review to search criteria

In this section we review the search criteria, which has been developed for $k = 1$. The off-diagonal elements of SPM are the SP, which measure the ability of a design in searching the true nonzero effect, β_{20}, from an alternative $\beta_{2i} \in \beta_2$. That is, the probability of discriminating the true model, M_0, from another candidate model. Shirakura et al. (1996) considered the least discrimination strength of a design and suggested to take the smallest off-diagonal elements of SPM as a search capability criteria of a search design. That is,

$$P(\rho) = \min_{\beta_{20}} \min_{\mathcal{A}(\beta_{20};\beta_2)} G(\beta_{20}, \beta_{2i}, \rho). \tag{5}$$

This criterion depends on the unknown parameter ρ. So, using $P(\rho)$ for comparing two search designs, at a given ρ-value, provides an enormous task. To overcome this problem, Ghosh and Teschmacher (2002) considered $|\rho|$ and used the property $|x| < 1$ to show that G in (4) is an increasing function of c_1, at a given r_{20}. This provides a one-to-one correspondence between the elements in SPM and a matrix with elements c_1 (CM), which are given in (4). They proposed three criteria, based on comparing the elements of CMs of two rival search designs. That is, let CM_1-CM_2 be the difference of c_1 matrices for two designs, say T_1 and T_2. Let $\tilde{n}^+$ denotes the number of positive elements of CM_1-CM_2. If $\tilde{n}^+ > \nu_2(\nu_2 - 1)/2$, then T_1 is better than T_2 in searching the non zero effect $\beta_{20} \in \beta_2$. This procedure is independent of unknown parameter ρ and is able to compare two search designs once, however, in price of losing information in ρ. This may decrease the discrimination between two designs. So, the challenge would be for obtaining a more precise such a criterion. Talebi and Esmailzadeh (2009) gave a criterion based on SP, which is independent of ρ, with a higher degree of searching ability. They considered a weighted average of $G(\beta_{20}, \beta_{2i}, \rho)$ over ρ by a weight function $f(\rho)$. Namely,

$$W(\beta_{20}, \beta_{2i}) = \int_0^\infty G(\beta_{20}, \beta_{2i}, \rho) f(\rho) d\rho. \tag{6}$$

The weighted search probability WSP criterion for a design T, has been given by

$$P_T = \min_{\beta_{20}} \min_{\mathcal{A}(\beta_{20};\beta_2)} W(\beta_{20}, \beta_{2i}). \tag{7}$$

They took $f(\rho)$ to be the density function of a Gamma random variable $\omega = \rho^2$ with parameters υ and λ and mean $\frac{\upsilon}{\lambda}$. For this particular weight function,

$W(\beta_{20}, \beta_{2i})$ simplify to the following closed form expression:

$$W(\beta_{20}, \beta_{2i}) = 1 - \mathcal{T}\left(c_1(\frac{\upsilon}{\lambda})^{1/2}\right) - \mathcal{T}\left(c_2(\frac{\upsilon}{\lambda})^{1/2}\right) + 2\mathcal{T}\left(c_1(\frac{\upsilon}{\lambda})^{1/2}\right)\mathcal{T}\left(c_2(\frac{\upsilon}{\lambda})^{1/2}\right) \quad (8)$$

where $\mathcal{T}(.)$ is cdf of t-student random variable. This criterion enables one to compare two isomorphic class of search designs. It also works for heteroscedastic underlying search models. Note that all of above criteria are capable to do the evaluation and comparison search ability of search designs for $k = 1$. For general case $k \geq 1$, two new criteria have been developed, based on Kullback-Leibler distance by Talebi and Esmailzadeh (2010).

3. Optimal-search criterion

The criteria which have been reviewed in the previous section are so-called discrimination efficiency (DE) criteria of a design. By these criteria the estimation efficiency (EE) of a design is not taken into account. In contrast, many researchers have been presented criteria to obtain optimum design, which are efficient in parameter estimation. Most of these criteria are a function of covariance matrix of estimators. The most popular one is D-optimality, which minimize(maximize) the determinant of covariance (information) matrix. Several authors have developed criteria to do the dual task of model discrimination and parameters estimation. Mostly, these criteria have been given for nested regression models. In this section, we consider the partially non-nested models, given in (3), in the context of search linear models (1). In order to develop a dual task criterion for measuring the search and estimation efficiency of a design, we consider a convex weighted geometric average compound quantity by:

$$C = (DE)^{\alpha}(EE)^{1-\alpha}, \quad 0 < \alpha < 1. \quad (9)$$

We take the D-optimality as the EE part in C. However, note that there are $\binom{\nu_2}{k}$ candidate models, given in (3), with $N \times f$ models matrices $H_{2i} = (X_1; X_{2i})$, $i = 1, 2, ..., \binom{\nu_2}{k}$, where $f = \nu_1 + k$. The $f \times f$ information matrix for the i-th model is $H'_{2i}H_{2i}$. Let define EE to be:

$$EE = \min_i \left[\frac{1}{N}|H'_{2i}H_{2i}|^{1/f}\right]. \quad (10)$$

Clearly, EE in (10) is a positive value, which is less than 1. For the case $k = 1$, we take DE to be P_T, given in (7). So, the positive quantity in (9), for the chosen DE and EE, is less than 1, i.e. $0 < C < 1$. Note that for large positive effect size, ρ, almost all rival designs are efficient in searching the true nonzero effects β_{20}. So, one may be interested in focussing on estimation efficiency of a design for large effect sizes. Therefore, by increasing property of P_T on ρ, through the hyper parameter υ and λ in $f(\rho)$, we suggest to take α to be a decreasing function of positive ρ. Let

we assume $\alpha = e^{-\theta\rho}$, where positive constant θ depends on balancedness priority of the DE and EE in (9). So, we propose the dual-task criterion by

$$C = [P_T]^{e^{-\theta\rho}} \left[\min_i \left(\frac{1}{N}|H'_{2i}H_{2i}|^{1/f}\right)\right]^{1-e^{-\theta\rho}}, \theta > 0,\ 0 < \rho < a, \qquad (11)$$

at a given level of ρ. One may calculate and use the mean efficiency, over a reasonable range $[0, a]$ on ρ, by:

$$C_m = \frac{1}{a}\int_0^a C d\rho. \qquad (12)$$

For $k \geq 1$, one may replace KL-criterion, given by Talebi and Esmailzadeh (2010) for DE. For more efficient design, in a sense of both search and estimation, a large value of C or C_m, i.e. closer to 1, is desirable. We will imply this criterion to compare some search designs in the next section.

Table 1. D1, D2, D3 and D4 with 12 runs and 4 factors

D1				D2				D3				D4			
A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
1	1	1	1	1	-1	1	-1	-1	1	-1	-1	1	1	1	1
-1	-1	-1	-1	1	1	-1	1	1	1	1	1	1	-1	-1	-1
-1	-1	-1	1	-1	1	1	-1	-1	-1	-1	-1	-1	1	-1	-1
-1	-1	1	-1	1	-1	1	1	-1	-1	1	-1	-1	-1	1	-1
-1	1	-1	-1	1	1	-1	1	-1	1	-1	-1	-1	-1	-1	1
1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	-1	-1	-1	-1
-1	-1	1	1	-1	1	1	1	-1	-1	1	1	-1	-1	1	1
-1	1	-1	1	-1	-1	1	1	-1	1	1	1	-1	1	1	1
1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1
-1	1	1	-1	1	-1	-1	-1	1	1	1	-1	1	1	-1	-1
1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	1	-1	-1	1
1	1	-1	-1	-1	-1	-1	-1	1	1	-1	1	-1	-1	-1	-1

4. Implementation

Consider search designs $D_1 - D_4$, given in Table 1, for a 2^4 factorial experiment. Designs $D_1 - D_3$ are given in Ghosh and Teschmacher (2002) and D_4 obtained from Latin square, given by Talebi and Esmailzadeh (2009). Design D_1 is a balanced array of full strength, D_2 is an orthogonal array and D_3 is partially orthogonal. These designs have been ranked with respect to their search performance, using WSP given in (7). The calculated WSP for some values of hyper parameters υ and λ are given in Table 2. The result is $P_{D_1} > P_{D_2} > P_{D_4} > P_{D_3}$. That is, D_1 is the most superior to the D_2, D_3 and D_4. Now, we use (11) and (12) to evaluate and rank the four designs with respect to both search and estimation capability. To do

Table 2. WSP values for D_1, D_2 , D_3 and D_4

(v, λ)	(1,1)	(1,2)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,4)	(4,2)
D_1	0.8381	0.9020	0.9516	0.9822	0.9908	0.9837	0.9962	0.9993	0.9991
D_2	0.8315	0.8976	0.9489	0.9813	0.9904	0.9829	0.9961	0.9993	0.9991
D_3	0.7735	0.8515	0.9057	0.9597	0.9776	0.9576	0.9877	0.9973	0.9959
D_4	0.7886	0.8632	0.9152	0.9641	0.9801	0.9622	0.9891	0.9976	0.9964

this, we take $\theta = 0.5$ and $a = 5$. Note that $N = 12$ and $f = 6$. Figure 1 presents the graphs of the C in (11) for D_1, D_2, D_3 and D_4 as functions of ρ. It shows that D_1 and D_2 are better than D_3 and D_4 for all values of $\rho > 0$ and D_4 is better than D_3, however D_2 is inferior to D_1 only for lower values of ρ. It means that the comparison between D_1 and D_2 is inconclusive for all ρ, if one uses the C criterion in (11). The mean efficiency, C_m given in (12) solves this inconclusive problem. The values of C_m, for four designs, are given in Table 3. The result shows that based on this criterion D_2 is superior to D_1, D_3 and D_4, D_1 is better than D_4 and D_4 is better than D_3.

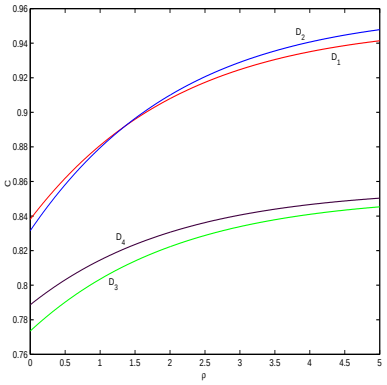

Fig. 1. Plot of C for D_1, D_2, D_3 and D_4

Table 3. C_m values for D_1, D_2 , D_3 and D_4

(v, λ)	(1,1)	(1,2)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,4)	(4,2)
D_1	0.9085	0.9329	0.9513	0.9625	0.9656	0.9630	0.9676	0.9687	0.9686
D_2	0.9106	0.9361	0.9553	0.9671	0.9704	0.9677	0.9725	0.9736	0.9736
D_3	0.8226	0.8519	0.8715	0.8905	0.8967	0.8898	0.9002	0.9035	0.9030
D_4	0.8309	0.8587	0.8775	0.8947	0.9002	0.8940	0.9033	0.9063	0.9058

5. Conclusion

A design, which is more efficient in a sense of model discrimination may not be good in parameter estimation of the underlying models. An example is D_1 and D_2 designs in this paper. The search superiority of D_1 to D_2 has been shown by Ghosh and Teschmacher (2002) and confirmed by the WSP given in (7) too. The later is presented in Table 2. However, the values of C_m-criterion, given in Table 3, show that D_1 is inferior to D_2. That is, when both discrimination and estimation efficiency of search designs have been taken into account once, the conclusion may reverse. Of course, this may depend on the chosen criterion and subjective priority of design experimenter, for balanced made between two efficiency properties. The criteria, given in this paper are very flexible in a sense of making balance priority between efficiency properties.

References

1. Atkinson,A.C., (2008), DT-optimum designs for model discrimination and parameter estimation. J. Statist. Plann. Inference 138, 56-64
2. Biswass, A. and Chaudhuri, p., (2002), An efficient design for model discrimination and parameter estimation in linear models. Biometrika, 89, 709-718.
3. Dette, H., (1993). On a mixture of the D- and D1-optimality criterion in polynomial regression. J. Statist. Plann. Inference 35, 233-249.
4. Ghosh, S. and Talebi, H., (1993), Main effect plans with additional search property for factorial experiments, J. Statist. Plann. Inference, 36, 367-384.
5. Ghosh,S. , (1996) Sequential assembly of fractions in factorial experiments. Handbook of Statistics, Vol.13, 407-435.
6. Ghosh,S. , Teschmacher,L., (2002), Comparisons of search designs using search probabilities. J. Statist. Plann. Inference 104, 439-458.
7. Gupta, B. C. (1984), A lower bound for number of treatments in a main effect plus one plan for 2^4 factorials. Utilitas Math. 26, 259-267.
8. Gupta, B. C. (1988), A bound connected with 2^6 factorial search designs of Resolution 3.1. Comm. Statist. Theory Methods 17, 3137-3144.
9. Ohnishi, T. and T. Shirakura (1985), Search designs for 2m factorial experiments. J. Statist. Plann. lnJerence 11, 241-245.
10. Shirakura, T., 1991. Main effect plus one or two plans for 2m factorials, J. Statist. Plann. Inference 27, 65-74.
11. Shirakura, T., Takahashi, T. and Srivastava, J.N., (1996), Searching probabilities for nonzero effects in search designs for the noisy case. Ann. Statist. 24 (6), 2560-2568.
12. Srivastava, J.N., (1975), Designs for searching non-negligible effects. In: Srivastava, J.N. (Ed.), A Survey of Statistical Design and Linear Models. North-Holland, Elsevier, Amsterdam, pp. 507-519.
13. Talebi, H. and Esmailzadeh, N., (2009), Weighted searching probability for classes of equivalent search designs Comparison. Comm. Statist. Theory Methods (to appear).
14. Talebi, H. and Esmailzadeh, N., (2010), Using Kullback-Leibler distance for performance evaluation of search designs. (submitted)
15. Tommasi, C., (2009), Optimal designs for both model discrimination and parameter estimation. J. Statist. Plann. Inference 139, 4123-4132.

A Note on Stochastic Comparisons and Aging Properties of Conditional Generalized Order Statistics

M. Tavangar and M. Asadi

Department of Statistics, University of Isfahan

The generalized order statistics (GOS) model is a unified model which contains the well known ordered random data such as order statistics and record values. In the present paper, we study some partial ordering results and aging properties of the conditional GOS. The results of the paper subsume some of the existing results which recently are obtained in the literature. In particular, our results hold for the model of progressively Type-II right censored order statistics without any restriction on the censoring scheme.

Keywords: Likelihood ratio order; Residual lifetime; Inactivity time; Mean residual life; Progressive censoring.

1. Introduction

In the past ten years, several results have been appeared in the literature which deal with the stochastic comparisons of the residual lifetime or inactivity time of order statistics and record values . Some recent articles on the subject are Franco *et al.* (2001), Belzunce *et al.* (2005), Hu and Zhuang (2005), Xie and Hu (2008), Asadi and Bayramoglu (2005,2006), Li and Zhao (2006,2008), Asadi (2006), Khaledi and Shaked (2007), Khaledi and Shojaei (2007), Raqab and Asadi (2008) and Tavangar and Asadi (2009a,2009b). Hu *et al.* (2007) and Zhao and Balakrishnan (2009) have generalized some of these results to generalized order statistics (GOS) with equal parameters. In this paper, we generalize some of the existing results in the literature to compare the residual lifetime or the inactivity time of GOS under less restrictions on the parameters of the GOS model. We also provide some results on the aging properties of conditional GOS. It is shown that our findings cover various useful models of ordered random variables, mentioned in Kamps (1995), such as the sequential order statistics and Pfeifer record model. Some interesting comparisons on the residual lifetimes and the inactivity times of progressively Type-II right censored order statistics with an arbitrary censoring scheme $\tilde{R}$ are also established.

The concept of GOS is introduced by Kamps (1995). He showed that order statistics, record values, progressively Type-II right censored order statistics and some other ordered random variables can be considered as special cases of the

GOS model. For more details on properties of these models, readers may refer to Kamps (1995), Arnold *et al.* (1992,1998), David and Nagaraja (2003) and Balakrishnan and Aggarwala (2000). Let $F(x)$ be an absolutely continuous distribution function with density function $f(x)$. Let also $\bar{F}(x) = 1 - F(x)$ denote the survival function. The random variables $X(1, n, \widetilde{m}, k), X(2, n, \widetilde{m}, k), \ldots, X(n, n, \widetilde{m}, k)$ are called the GOS based on F if their joint density function is given by

$$f_{X(1,n,\widetilde{m},k),\ldots,X(n,n,\widetilde{m},k)}(x_1, \ldots, x_n) = k \left[\prod_{j=1}^{n-1} \gamma_j\right] \left[\prod_{i=1}^{n-1} \{\bar{F}(x_i)\}^{m_i} f(x_i)\right] \{\bar{F}(x_n)\}^{k-1} f(x_n),$$
$$F^{-1}(0) < x_1 \leq \cdots \leq x_n < F^{-1}(1-),$$

where $n \in \mathbb{N}$, $k > 0$ and $m_1, m_2, \ldots, m_{n-1} \in \mathbb{R}$ such that $\gamma_r = k + n - r + \sum_{j=r}^{n-1} m_j \geq 1$ for all $r \in \{1, 2, \ldots, n-1\}$, and $\widetilde{m} = (m_1, m_2, \ldots, m_{n-1})$, if $n \geq 2$ ($\widetilde{m} \in \mathbb{R}$ is arbitrary, if $n = 1$). Based on the different values of the parameters of the model one can obtain the joint density of order statistics, record values, etc.

Before giving the main results, we first recall some stochastic orders that are pertinent to the developments of the paper.

Definition 1.1. For two random variables X and Y, with their densities f and g and distribution functions F and G, respectively, let $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$ be their survival functions. As the ratios in the statements below are well defined, X is said to be smaller than Y in

(a) likelihood ratio order (denoted by $X \leq_{\rm lr} Y$) if $g(x)/f(x)$ is increasing in x;
(b) hazard rate order (denoted by $X \leq_{\rm hr} Y$) if $\bar{G}(x)/\bar{F}(x)$ is increasing in x;
(c) reversed hazard rate order (denoted by $X \leq_{\rm rh} Y$) if $G(x)/F(x)$ is increasing in x;
(d) stochastic order (denoted by $X \leq_{\rm st} Y$) if $\bar{G}(x) \geq \bar{F}(x)$.

For a comprehensive discussion on these stochastic orders, we refer the reader to Shaked and Shanthikumar (2007) and Müller and Stoyan (2002).

We recall that a non-negative function $h(x, y)$ is said to be totally positive of order 2 ($\rm TP_2$) if

$$h(x_1, y_1)h(x_2, y_2) - h(x_1, y_2)h(x_2, y_1) \geq 0,$$

whenever $x_1 < x_2$ and $y_1 < y_2$. Karlin (1968) provides some important properties of $\rm TP_2$ functions.

The next lemma will be helpful in deriving our main results.

Lemma 1.1. *(Misra and van der Meulen, 2003). Assume that Θ is a subset of the real line $\mathbb{R}$, and U is a non-negative random variable having a distribution function belonging to the family $\mathcal{P} = \{H(. \mid \theta), \theta \in \Theta\}$, which is such that, for $\theta_1, \theta_2 \in \Theta$, and $\theta_1 < \theta_2$,*

$$\bar{H}(x \mid \theta_1) \leq (\geq) \bar{H}(x \mid \theta_2), \quad x \geq 0.$$

Let $\psi(u,\theta)$ be a real valued function defined on $\mathbb{R}\times\Theta$, which is measurable in u for each θ such that $E_\theta\{\psi(U,\theta)\}$ exists. Then

(a) *$E_\theta\{\psi(U,\theta)\}$ is increasing in θ, if $\psi(u,\theta)$ is increasing in θ and increasing (decreasing) in u; and*

(b) *$E_\theta\{\psi(U,\theta)\}$ is decreasing in θ, if $\psi(u,\theta)$ is decreasing in θ and decreasing (increasing) in u.*

Throughout the paper, for any random variable W, f_W and $\bar{F}_W$ denote its probability density function and survival function, respectively.

2. Stochastic comparisons

In the following, we first provide a modified version of the result of Hu and Zhuang (2005) which will be used in next section. For our derivations, we consider a parameter vector $\tilde{\mu}$ which is different from that used in Hu and Zhuang (2005). For proof see Tavangar and Asadi (2010).

Theorem 2.1. *Let $X(1,n,\tilde{m},k), X(2,n,\tilde{m},k),\ldots,X(n,n,\tilde{m},k)$ be the GOS based on a continuous distribution function $F(x)$ and $\tilde{\mu}=(m_2,m_3,\ldots,m_{n-1})$, $n\in\mathbb{N}$. Then*

(a) *$X(r-1,n,\tilde{m},k)\leq_* X(r,n,\tilde{m},k)$, $r=2,3,\ldots,n$;*

(b) *$X(r,n,\tilde{m},k)\leq_* X(r,n-1,\tilde{\mu},k)$, $r=1,2,\ldots,n-1$, if $m_i\geq -1$ for each $i=1,2,...,r$, and*

(c) *$X(r-1,n-1,\tilde{\mu},k)\leq_* X(r,n,\tilde{m},k)$, $r=2,3,\ldots,m$,*

where the order "$\leq_*$" is "$\leq_{\rm lr}$" if F is absolutely continuous, and is "$\leq_{\rm hr}$" or "$\leq_{\rm rh}$" if F is continuous.

It should be mentioned here that part (a) of Theorem 2.1 is proved in Hu and Zhuang (2005) under the assumption that $m_j\geq -1$ for each j. We will use this theorem in Section 3 to study some aging and monotonicity properties of the conditional GOS.

Assume that X and Y are two absolutely continuous random variables with the distribution functions F and G, the density functions f and g, and the hazard rates λ_1 and λ_2, respectively. Further, let $X(r,n,\tilde{m},k)$ and $Y(r,n,\tilde{m}',k')$ denote the GOS based on F and G, with parameters k and m_i, and k' and m_i', $i=1,2,...,n-1$, respectively. Tavangar and Asadi (2010) proved the following theorem which is an extension of a result of Zhao and Balakrishnan (2009) under the more general assumptions on the parameters of the model of GOS.

Theorem 2.2. *Let $k\geq k'$ and $m_i\geq m_i'$ for all $i=1,2,...,n-1$. For $1\leq r\leq s\leq n$ and any $t\in\mathbb{R}$, suppose that (a) $m_i\geq 0$ or $m_i'\geq 0$ for all $i=1,2,...,n-1$ and $X\leq_{\rm lr} Y$, or (b) $m_i\geq -1$ or $m_i'\geq -1$ for all $i=1,2,...,n-1$, $X\leq_{\rm hr} Y$ and $\lambda_2(x)/\lambda_1(x)$ is increasing in x. Then*

(i) *$[X(s,n,\tilde{m},k)-t\mid X(r,n,\tilde{m},k)>t]\leq_{\rm lr}[Y(s,n,\tilde{m}',k')-t\mid Y(r,n,\tilde{m}',k')>t]$;*

(ii) $[t - X(r, n, \tilde{m}, k) \mid X(s, n, \tilde{m}, k) \leq t] \geq_{\rm lr} [t - Y(r, n, \tilde{m}', k') \mid Y(s, n, \tilde{m}', k') \leq t]$.

3. Aging properties

In this section, we obtain some stochastic properties of conditional GOS

$$[X(s, n, \tilde{m}, k) - t \mid X(r, n, \tilde{m}, k) > t], \quad 1 \leq r \leq s \leq n, \tag{1}$$

which, in recent years, has aroused the interest of many authors. For some discussions along this line, one may refer to Asadi and Bayramoglu (2005, 2006), Li and Zhao (2006), Khaledi and Shaked (2007), Hu *et al.* (2007) and Zhao and Balakrishnan (2009). To prove the results of this section, we apply the following useful representation for the survival function of (1) given in Tavangar and Asadi (2010):

$$\bar{H}^{\tilde{m}}_{r,s,n,t}(x) = \frac{\sum_{j=0}^{r-1} \gamma_{j+1}^{-1} f_{X(j+1,n,\tilde{m},k)}(t) \bar{F}_{X_t(s-j,n-j,\tilde{\mu}_j,k)}(x)}{\sum_{j=0}^{r-1} \gamma_{j+1}^{-1} f_{X(j+1,n,\tilde{m},k)}(t)}. \tag{2}$$

Here $X_t = (X - t \mid X > t)$ denotes the residual life random variable. Representation (2) can be used to establish the properties of conditional GOS from those of unconditional ones.

In reliability theory, to describe the aging properties of lifetime random variables, there have been defined several aging concepts. The best studied (univariate) ageing classes are increasing failure rate (IFR) and decreasing failure rate (DFR) classes. A lifetime distribution with survival function $\bar{F}$ belongs to IFR (DFR) class if $\bar{F}(x + t)/\bar{F}(t)$ is decreasing (increasing) in t for each $x > 0$. Distributions with a Lebesgue density belong to the IFR (DFR) class if and only if their hazard rates $\lambda(t) = f(t)/\bar{F}(t)$ are increasing (decreasing) (see Barlow and Proschan, 1978). Lifetimes with IFR distributions occur in many situations. Examples are failure times of diverse mechanical units and lifetimes of humans after some initial period.

We will now focus on stochastic monotone properties of $[X(s, n, \widetilde{m}, k) - t \mid X(r, n, \widetilde{m}, k) > t]$ with respect to t. The following lemma will be useful in proving next result.

Lemma 3.1. *For any integer r such that $1 \leq r \leq n$, $[X(r, n, \widetilde{m}, k) - t \mid X(1, n, \widetilde{m}, k) > t]$ is stochastically decreasing (increasing) in $t \in \mathbb{R}_+$ if and only if X is IFR (DFR).*

Theorem 3.1. *For any two integers r and s such that $1 \leq r \leq s \leq n$, and arbitrary $\tilde{m} \in \mathbb{R}^{n-1}$,*

(a) *if X is IFR, then $[X(s, n, \widetilde{m}, k) - t \mid X(r, n, \widetilde{m}, k) > t]$ is stochastically decreasing in $t \in \mathbb{R}_+$;*

(b) *if $[X(s, n, \widetilde{m}, k) - t \mid X(r, n, \widetilde{m}, k) > t]$ is stochastically increasing in $t \in \mathbb{R}_+$, then X is DFR.*

Proof. First note that by Representation (2), the survival function $\bar{H}^{\tilde{m}}_{r,s,n,t}(x)$ can be rewritten as

$$\bar{H}^{\tilde{m}}_{r,s,n,t}(x) = E_t\{\psi(U,t)\},$$

where $\psi(j,t) = \bar{H}^{\tilde{\mu}_j}_{1,s-j,n-j,t}(x)$ and the distribution function of the random variable U belongs to the family $\mathcal{P} = \{H(.\mid t), t \in \mathbb{R}_+\}$ with densities

$$h(j \mid t) = \frac{\gamma_{j+1}^{-1} f_{X(j+1,n,\tilde{m},k)}(t)}{\sum_{i=0}^{r-1} \gamma_{j+1}^{-1} f_{X(j+1,n,\tilde{m},k)}(t)}, \quad j = 0, 1, ..., r-1.$$

(a) Let X be IFR. Then by Lemma 3.1, $\psi(j,t)$ is decreasing in $t \in \mathbb{R}_+$ for all j. From part (c) of Theorem 2.1, we have

$$X_t(s-j-1, n-j-1, \tilde{\mu}_{j+1}, k) \leq_{\rm st} X_t(s-j, n-j, \tilde{\mu}_j, k).$$

This, in turn, implies that $\psi(j+1,t) \leq \psi(j,t)$ and hence $\psi(j,t)$ is decreasing in j for all $t \in \mathbb{R}_+$. It follows from part (a) of Theorem 2.1 that $h(j+1 \mid t)/h(j \mid t)$ is increasing in $t \in \mathbb{R}_+$. Thus $h(j \mid t)$ is TP_2 in $(j,t) \in \{0,1,...,r-1\} \times \mathbb{R}_+$ which, in turn, implies that $\bar{H}(j \mid t_1) \leq \bar{H}(j \mid t_2)$, $j = 0, 1, ..., r-1$, whenever $t_1 < t_2$. Taking into account these observations, from Lemma 1.1, we can conclude that $E_t\{\psi(U,t)\}$ is a decreasing function of t, from which the result of part (a) of the theorem follows.

(b) If $E_t\{\psi(U,t)\}$ is increasing in $t \in \mathbb{R}_+$, then we must have $\psi(j,t)$ is increasing in $t \in \mathbb{R}_+$ for some $j \in \{0,1,...,r-1\}$. Hence there exists some $j \in \{0,1,...,r-1\}$ such that $\bar{H}^{\tilde{\mu}_j}_{1,s-j,n-j,t}(x)$ is increasing in $t \in \mathbb{R}_+$ which , by Lemma 3.1, in turn implies that X is DFR. The proof of the theorem is complete. □

The next result carries out the stochastic comparisons of conditional GOS and unconditional ones.

Theorem 3.2. *For any two integers r and s such that $1 \leq r \leq s \leq n$, and arbitrary $\tilde{m} \in \mathbb{R}^{n-1}$, if X is NBU, then*

$$[X(s,n,\widetilde{m},k) - t \mid X(r,n,\widetilde{m},k) > t] \leq_{\rm st} X(s,n,\widetilde{m},k),$$

for $t \in \mathbb{R}_+$, and if X is NWU, then

$$[X(s,n,\widetilde{m},k) - t \mid X(r,n,\widetilde{m},k) > t] \geq_{\rm st} X(s-r+1, n-r+1, \widetilde{\mu}_{r-1}, k),$$

for $t \in \mathbb{R}_+$.

Proof. For $1 \leq r \leq n$, let $\bar{H}^{\tilde{m}}_{r,n}(x)$ denote the survival function of $X(r,n,\widetilde{m},k)$. If X is NBU (NWU), then $F_t(x) \geq (\leq) F(x)$ and hence

$$\bar{H}^{\tilde{m}}_{1,s,n,t}(x) = P\{U(s,n,\widetilde{m},k) > F_t(x)\} \leq (\geq) P\{U(s,n,\widetilde{m},k) > F(x)\} = \bar{H}^{\tilde{m}}_{s,n}(x),$$

where $F_t(x)$ denotes the distribution function of the residual life random variable. According to Theorem 2.1(c) we have

$$\bar{H}^{\tilde{\mu}_{r-1}}_{1,s-r+1,n-r+1,t}(x) \leq \bar{H}^{\tilde{\mu}_j}_{1,s-j,n-j,t}(x) \leq \bar{H}^{\tilde{m}}_{1,s,n,t}(x),$$

which implies that

$$\bar{H}^{\tilde{\mu}_{r-1}}_{1,s-r+1,n-r+1,t}(x) \leq \bar{H}^{\tilde{m}}_{r,s,n,t}(x) \leq \bar{H}^{\tilde{m}}_{1,s,n,t}(x).$$

By combining these results, we obtain $\bar{H}^{\tilde{m}}_{r,s,n,t}(x) \leq \bar{H}^{\tilde{m}}_{s,n}(x)$ when X is NBU, and $\bar{H}^{\tilde{m}}_{r,s,n,t}(x) \geq \bar{H}^{\tilde{\mu}_{r-1}}_{s-r+1,n-r+1}(x)$ when X is NWU. Hence, the desired result follows. □

Remark 3.1. Some well known IFR distributions are the Gamma distribution and the Weibull distribution both with shape parameters greater than one, the half-normal distribution, and the half-logistic distribution. It is known that the class of IFR distribution is contained in the class of NBU distribution (see, for example, Barlow and Proschan, 1975). Hence the cited distributions are also NBU. There exists a similar relation between DFR and NWU distributions. Some DFR statistical models are the Gamma and the Weibull distributions with shape parameters less than one, the Pareto distribution with shape parameter greater than one, and the mixture of two exponential distributions.

4. Applications in progressive censoring

An important application of the results established in this section is in the field of progressive censoring. Let $X_1, X_2, \ldots, X_n$ denote the failure times of n independent and identically distributed (i.i.d.) items which are placed on a life-test. Suppose that $R_1, R_2, \ldots, R_m$ are some fixed non-negative integers such that $\sum_{v=1}^{m} R_v = n - m$. It is planned that only m failures will be observed and the remaining $n - m$ lifetimes will be censored progressively according to the censoring scheme $\tilde{R} = (R_1, R_2, \ldots, R_m)$. More specifically, at the time of the ith failure, R_i surviving items will be randomly withdrawn from the experiment, $i = 1, 2, \ldots, m$. The resulting ordered observed failure times, denoted by $X^{\tilde{R}}_{1:m:n}, X^{\tilde{R}}_{2:m:n}, \ldots, X^{\tilde{R}}_{m:m:n}$, are called the progressively Type-II right censored order statistics of size m from a sample of size n with a progressive censoring scheme $\tilde{R}$. These ordered random variables form a special case of GOS if we set $n = m$, $m_i = R_i,\ i = 1, 2, ..., m-1$, and $k = R_m + 1$ (in the model of GOS). We refer the reader to Balakrishnan and Aggarwala (2000) and references therein for a comprehensive discussion and inferential procedures based on progressive censoring.

Let $X^{\tilde{R}}_{1:m:n}, X^{\tilde{R}}_{2:m:n}, \ldots, X^{\tilde{R}}_{m:m:n}$ be m progressively Type-II right censored order statistics arising from n i.i.d. non-negative random variables distributed as X, and similarly let $Y^{\tilde{S}}_{1:m:N}, Y^{\tilde{S}}_{2:m:N}, \ldots, Y^{\tilde{S}}_{m:m:N}$ be m progressively Type-II right censored order statistics arising from n i.i.d. non-negative random variables distributed as Y. From Theorem 2.2 we obtain the following corollary which gives some comparisons

of the residual lifetimes and the inactivity times of progressively Type-II right censored order statistics.

Corollary 4.1. *Let* $R_i \geq S_i$ *for all* $i = 1, 2, ..., m-1$ *and* $X \leq_{\rm lr} Y$. *Then for* $1 \leq r \leq s \leq m \leq \min\{n, N\}$ *and* $t \in \mathbb{R}_+$,

(a) $[X^{\tilde{R}}_{s:m:n} - t \mid X^{\tilde{R}}_{r:m:n} > t] \leq_{\rm lr} [Y^{\tilde{S}}_{s:m:N} - t \mid Y^{\tilde{S}}_{r:m:N} > t]$, *and*
(b) $[t - X^{\tilde{R}}_{r:m:n} \mid X^{\tilde{R}}_{s:m:n} \leq t] \geq_{\rm lr} [t - Y^{\tilde{S}}_{r:m:N} \mid Y^{\tilde{S}}_{s:m:N} \leq t]$.

In recent years, various concepts of the mean residual life and the mean inactivity time (or the mean past life) of ordered data have been defined. Bairamov *et al.* (2002) and Asadi and Bayramoglu (2005,2006) studied the mean residual life of k-out-of-n systems. Asadi (2006) and Tavangar and Asadi (2009b) obtained some properties of the mean inactivity time of k-out-of-n systems. The mean residual life of record values is studied in Raqab and Asadi (2008). Recently, Hashemi *et al.* (2010) investigated the mean residual life

$$E\left(X^{\tilde{R}}_{s:m:n} - t \mid X^{\tilde{R}}_{r:m:n} \leq t < X^{\tilde{R}}_{r+1:m:n}\right),$$

of progressively Type-II right censored order statistics given that exactly r items, in the experiment, failed at time or before time t.

Consider an experiment with arbitrary censoring scheme $\tilde{R}$. For design engineers and survival analysts, it will be of interest to have a knowledge of the properties of the mean residual life of unfailed items in the experiment. For this reason, one can consider

$$M^{\tilde{R}}_{r,s,m}(t) = E\left(X^{\tilde{R}}_{s:m:n} - t \mid X^{\tilde{R}}_{r:m:n} \geq t\right), \quad 1 \leq r \leq s \leq m \leq n,$$

the mean residual life of the sth failure time given that at most $(r-1)$ failures occurred at or before time t.

Now, we can apply Theorems 3.1 and 3.2 to obtain the following corollary regarding the mean residual life $M^{\tilde{R}}_{r,s,m}(t)$ of progressively Type-II right censored order statistics defined above.

Corollary 4.2. *For any two integers* r *and* s *such that* $1 \leq r \leq s \leq m \leq n$, *and arbitrary censoring scheme* $\tilde{R}$, *we have the following results:*

(a) *If* F *is IFR, then* $M^{\tilde{R}}_{r,s,m}(t)$ *is decreasing in* $t \in \mathbb{R}_+$;
(b) *If* F *is NBU, then*

$$M^{\tilde{R}}_{r,s,m}(t) \leq E\left(X^{\tilde{R}}_{s:m:n}\right),$$

and if F *is NWU, then*

$$M^{\tilde{R}}_{r,s,m}(t) \geq E\left(X^{(R_r, R_{r+1}, ..., R_m)}_{s-r+1:m-r+1:\gamma_r}\right).$$

By using the results of this paper, one can also establish some stochastic orderings and aging monotonicity results for other models of ordered random variables such as Pfeifer records and sequential order statistics based on general distributions $F_1, F_2, \ldots, F_n$.

References

1. Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1992). *A First Course in Order Statistics*. Wiley, New York.
2. Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1998). *Records*. Wiley, New York.
3. Asadi, M. (2006). On the mean past lifetime of the components of a parallel system. *Journal of Statistical Planning and Inference*, **136**, 1197–1206.
4. Asadi, M. and Bayramoglu, I. (2005). A note on the mean residual life function of a parallel system. *Communications in Statistics: Theory and Methods*, **34**, 475–484.
5. Asadi, M. and Bayramoglu, I. (2006). On the mean residual life function of the k-out-of-n systems at the system level. *IEEE Transactions on Reliability*, **55**, 314–318.
6. Bairamov, I., Ahsanullah, M. and Akhundov, I. (2002). A residual life function of a system having parallel or series structures. *Journal of Statistical Theory and Applications*, **1** (2), 119–132.
7. Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods, and Applications*. Birkhauser, Boston.
8. Barlow, R.E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart: Winston, New York.
9. Belzunce, F., Mercader, J.A. and Ruiz, J.M. (2005). Stochastic comparisons of generalized order statistics. *Probability in the Engineering and Informational Sciences*, **19**, 99–120.
10. David, H.A. and Nagaraja, H.N. (2003). Order Statistics. Third edition, Wiley, New Jersey.
11. Franco, M., Ruiz, J.M. and Ruiz, M.C. (2001). Stochastic orderings between spacings of generalized order statistics. *Probability in the Engineering and Informational Sciences*, **16**, 471–484.
12. Hashemi, M., Tavangar, M. and Asadi, M. (2010). Some properties of the residual lifetime of progressively Type-II right censored order statistics. To appear in: *Statistics and Probability Letters*.
13. Hu, T., Jin, W. and Khaledi, B.-E. (2007). Ordering conditional distributions of generalized order statistics. *Probability in the Engineering and Informational Sciences*, **21**, 401–417.
14. Hu, T. and Zhuang, W. (2005). A note on comparisons of generalized order statistics. *Statistics and Probability Letters*, **72**, 163–170.
15. Kamps, U. (1995). *A concept of generalized order statistics*. B.G. Teubner, Stuttgart.
16. Karlin, S. (1968). *Total Positivity*. Stanford University Press, Palo Alto, California.
17. Khaledi, B.-E. and Shaked, M. (2007). Ordering conditional lifetimes of coherent systems. *Journal of Statistical Planning and Inference*, **137**, 1173–1184.
18. Khaledi, B.-E. and Shojaei, R. (2007). On stochastic ordering between residual record values. *Statistics and Probability Letters*, **77**, 1467–1472.
19. Li, X. and Zhao, P. (2006). Some aging properties of the residual life of k-out-of-n

systems. *IEEE Transaction on Reliability*, **55** (3), 535–541.

20. Li, X. and Zhao, P. (2008). Stochastic comparison on general inactivity time and general residual life of k-out-of-n systems. *Communications in Statistics: Simulation and Computation*, **37**, 1005–1019.
21. Misra, N. and van der Meulen, E.C. (2003). On stochastic properties of m spacings. *Journal of Statistical Planning and Inference*, **115**, 683–697.
22. Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley, New York.
23. Raqab, M.Z. and Asadi, M. (2008). On the mean residual life of records. *Journal of Statistical Planning and Inference*, **138**, 3660–3666.
24. Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. Springer, New York.
25. Tavangar, M. and Asadi, M. (2009a). Some results on conditional expectations of lower record values. To appear in: *Statistics*.
26. Tavangar, M. and Asadi, M. (2009b). A study on the mean past lifetime of the components of $(n-k+1)$-out-of-n system at the system level. To appear in: *Metrika*.
27. Tavangar, M. and Asadi, M. (2010). Some aging properties of conditional generalized order statistics. *Submitted for publication*.
28. Xie, H. and Hu, T. (2008). Conditional ordering of generalized order statistics revisited. *Probability in the Engineering and Informational Sciences*, **22**, 334–346.
29. Zhao, P. and Balakrishnan, N. (2009). Stochastic comparisons and properties of conditional generalized order statistics. *Journal of Statistical Planning and Inference*, **139**, 2920–2932.

The Gamma-Uniform Distribution

H. Torabi and N. H. Montazeri

Department of Statistics, Yazd University

Up to present for modeling and analyzing of random phenomenons, some statistical distributions are proposed. This paper considers a new general class of distributions, generated from the logit of the gamma random variable. A special case of this family is the *Gamma-Uniform* distribution. We derive expressions for the four moments, variance, skewness, kurtosis, Shannon and Rényi entropy. We also discuss the asymptotic distribution of the extreme order statistics and simulation issues.

Keywords: Gamma-Uniform Distribution, Expintegrale Function, Regularized Incomplete Gamma Function, Confluent Hypergeometric Function, Shannon And RÉNyi Entropy, Pochhammer Symbol.

1. Introduction

In recent years a class of distribution was proposed based on logit of a beta random variable. One major benefit of the Beta generalized class of distributions is its ability of fitting skewed data that cannot be properly fitted by existing distributions. Starting from a parent cumulative distribution function (cdf) $G(x)$, Eugene et al. (2002) defined a class of generalized distributions by

$$F(x) = \frac{1}{B(\alpha,\beta)} \int_0^{G(x)} w^{\alpha-1}(1-w)^{\beta-1}\, dw = I_{G(x)}(\alpha,\beta), \qquad \alpha > 0,\ \beta > 0$$

where $I_z(a,b) = \frac{B_z(a,b)}{B(a,b)}$ is the regularized incomplete Beta function and $B_z(a,b) = \int_0^z t^{a-1}(1-t)^{b-1}\, dt, \quad 0 < z < 1$ is the incomplete Beta function and $B(a,b) = B_1(a,b)$ is the Euler Beta function.

Note that the supports of random variables associated with $F(.)$ and $G(.)$ are equal. This class of generalized distributions has been receiving considerable attention over the last years, in particular after the works of Eugene et al. (2002) and Jones (2004). Eugene et al. (2002) introduced what is known as the Beta-Normal (BN) distribution by taking G to be cdf of the Normal distribution. Some more properties of this distribution have been studied by Nadarajah and Gupta (2005). Nadarajah and Kotz (2004) introduced what is known as the Beta-Gumbel (BG) distribution by taking G to be cdf of Gumbel distribution. Nadarajah and Gupta (2004) introduced the Beta-Fréchet (BF) distribution by taking G to be

the Fréchet distribution. Further, Nadarajah and Kotz (2006) examined the Beta-Exponential (BE) distribution by taking G to be cdf of Exponential distribution. Alfred Akinsete et al. (2006) defined the Beta-Pareto (BP) distribution by taking G to be cdf of Pareto distribution. Lee et al. (2007) defined the Beta-Weibull (BW) distribution by taking G to be cdf of Weibull distribution. Fredy Castellares et al. (2009) introduced the Beta Log-Normal (BLN) distribution by taking G to be cdf of Log-Normal distribution. Barreto et al. (2009) introduced the Beta Generalized Exponential distribution (BGE) by taking G to be cdf of Generalized Exponential distribution. Pescim et al. (2010) proposed the Beta-Generalized Half-Normal distribution (BGHN) by taking G to be cdf of Generalized Half-Normal distribution.

In this paper we will introduce a new general class of distribution generated from the logit of the Gamma distribution. In this new class, similar to the previous class if G denote the cdf of a random variable, then cdf for a generalized class of distributions can be defined by

$$F(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\frac{G(x)}{\overline{G}(x)}} e^{-\frac{w}{\beta}} w^{\alpha-1}\, dw = 1 - Q(\alpha, \frac{G(x)}{\beta\overline{G}(x)}), \quad \alpha > 0,\ \beta > 0 \quad (1)$$

where $\overline{G}(x) = 1 - G(x)$ and $Q(a, z) = \frac{\Gamma(a,z)}{\Gamma(a)}$ is the regularized incomplete gamma function and $\Gamma(a, z) = \int_z^{\infty} t^{a-1}e^{-t}\, dt,\ z > 0$ is the incomplete gamma function and $\Gamma(a)$ is Euler gamma function.

Note that the support of random variables associated with $F(.)$ and $G(.)$ are equal. The probability density function (pdf) corresponding to Eq. (1) can be written as

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \frac{g(x)}{\overline{G}^2(x)} \exp\Big(-\frac{G(x)}{\beta\overline{G}(x)}\Big)\Big(\frac{G(x)}{\overline{G}(x)}\Big)^{\alpha-1}, \qquad \alpha > 0,\ \beta > 0$$

where $g(x) = \frac{dG(x)}{dx}$ is the parent density function. The pdf $f(x)$ will be most tractable when both functions $G(x)$ and $g(x)$ have simple analytic expressions. In this article, we study the case when $G(x)$ is the cdf of the Uniform distribution in (a, b), i.e, $U(a, b)$. In this case, the random variable X is said to be have the Gamma-Uniform distribution and denoted by $GU(\alpha, \beta, a, b)$.

2. The Gamma-Uniform Distribution

One of the simplest distribution in Statistics is the Uniform distribution. Thus, we motivated to introduce the Gamma-Uniform (GU) distribution by taking G in Eq. (1) to be cdf of a U(a,b). The cdf of GU distribution becomes

$$F(x) = \int_0^{\frac{x-a}{b-x}} \frac{e^{-\frac{w}{\beta}} w^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}}\, dw = 1 - Q(\alpha, \frac{x-a}{\beta(b-x)}), \qquad a < x < b \qquad (2)$$

The corresponding probability density function (pdf) and the hazard rate function associated with Eq. (2) are:

$$f(x) = \frac{(b-a)e^{-\frac{x-a}{\beta(b-x)}}(\frac{x-a}{b-x})^{\alpha-1}}{(b-x)^2\beta^\alpha\Gamma(\alpha)}, \qquad a < x < b \tag{3}$$

and

$$\lambda(x) = \frac{(b-a)e^{-\frac{x-a}{\beta(b-x)}}(\frac{x-a}{b-x})^{\alpha-1}}{(b-x)^2\beta^\alpha\Gamma(\alpha, \frac{x-a}{\beta(b-x)})} \qquad a < x < b \tag{4}$$

respectively. Figure 2 illustrates some of the possible shapes of pdf, cdf and hazard rate function of GU distributions for various values of α, β and $a = 0$, $b = 5$.

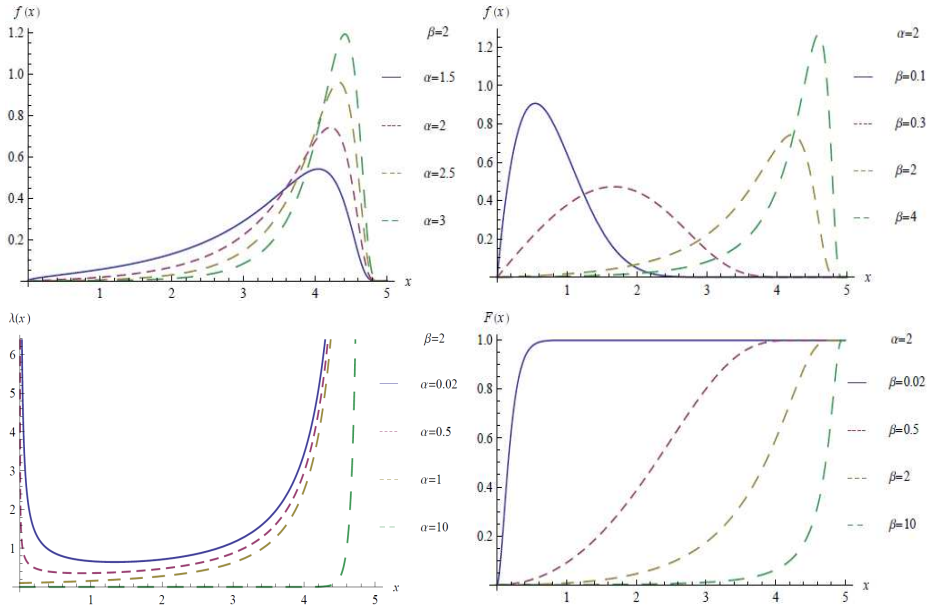

Fig. 1. The pdf, cdf and hazard rate function for $a = 0$ and $b = 5$.

3. Moments

In general, exact moments of a GU distribution cannot be calculated. However we derive some closed form expressions for the first four moments, variance, skewness and kurtosis.

In the subsequent, we will use the function $ExpIntegralE[n,z] = E[n,z]$ which is defined by $E[n,z] = \int_1^\infty e^{-\frac{zt}{t^n}}\ dt$.

$$
\begin{aligned}
E(X) &= \frac{1}{\beta}\big\{e^{\frac{1}{\beta}}\big(aE[\alpha,\frac{1}{\beta}] + b\alpha\beta E[1+\alpha,\frac{1}{\beta}]\big)\big\}\\
E(X^2) &= \frac{1}{\beta^2}\big\{\beta\big((a-b)^2+b^2\beta\big) - (a-b)e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]\big(a+a\beta(\alpha-1)-b(1+\beta+\alpha\beta)\big)\big\}\\
E(X^3) &= \frac{1}{2\beta^2}\Big\{\beta\Big(\alpha(a-b)^3+\beta\big(2a^3-3\alpha(a-b)^2(a+b)+\alpha^2(a-b)^3\big)\Big)\\
&\quad -(a-b)\alpha e^{\frac{1}{\beta}}E[1+\alpha,\frac{1}{\beta}]\Big(-2ab\big(1+\beta+2\alpha\beta+\beta^2(\alpha^2-1)\big)\\
&\quad +a^2\big(1+\beta(\alpha-1)(\beta(\alpha-2)+2)\big)+b^2\big(1+\beta(\alpha+2)(\beta+\alpha\beta+2)\big)\Big)\Big\}\\
E(X^4) &= \frac{1}{6\beta^4}\Big\{\beta\Big((a-b)^4+(a-b)^3\beta\big(a(2\alpha-3)-b(2\alpha+9)\big)+(a-b)^2\beta^2\\
&\quad \times\big(a^2(\alpha-3)(\alpha-2)-2ab(-6+\alpha+\alpha^2)+b^2(18+\alpha(\alpha+7))\big)+6b^4\beta^3\Big)\\
&\quad +(b-a)e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]\Big(-3a^2b\big(1+\beta+3\alpha\beta+\beta^2(\alpha-1)(2+3\alpha)\\
&\quad +\beta^3(\alpha-2)(\alpha-1)(\alpha+1)\big)+3ab^2(1+\beta+\alpha\beta)\big(1+\beta(\alpha+2)(2+\beta\\
&\quad \times(\alpha-1))\big)+a^3\big(1+(\alpha-1)\beta(3+\beta(\alpha-2)(3+\beta(\alpha-3)))\big)\\
&\quad +b^3\big(-1-\beta(\alpha+3)(3+(2+\alpha)\beta(3+\beta+\alpha\beta))\big)\Big)\Big\}
\end{aligned}
$$

Suppose X is a random variable with the pdf given by Eq. (3). If $a=0$ and $b=1$ then

$$E(X^n) = (\alpha)_n U\big(n, 1-\alpha, \frac{1}{\beta}\big)$$

where $(\alpha)_n = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}$ is Pochhammer symbol and $U(a,b,z) = \frac{1}{\Gamma(a)}\int_0^\infty e^{-zt}t^{a-1}(1+t)^{b-a-1}\ dt$ is confluent hypergeometric function.

Further calculations show that first three central moments, skewness and kurtosis of X can be given by

$$Var(X) = \frac{(a-b)^2}{\beta^2}\Big\{\beta - e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]\big(1+(\alpha-1)\beta+e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]\big)\Big\} =: \frac{(a-b)^2 s_1}{\beta^2}$$

$$E[\{X-E(X)\}^3] = \frac{(a-b)^3}{2\beta^3}\Big\{ -\beta(1+(\alpha-2)\beta) + e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]$$
$$\times\Big(1+2\beta(\alpha-4)+\beta^2(\alpha-2)(\alpha-1)+2e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]$$
$$\times\big(3+3\beta(\alpha-1)+2e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]\big)\Big)\Big\} =: \frac{(a-b)^3 s_2}{2\beta^3}$$

$$E[\{X-E(X)\}^4] = \frac{(a-b)^4}{6\beta^4}\Big\{\beta\Big(1+\beta\big(-3+2\alpha+(6-5\alpha+\alpha^2)\beta\big)\Big)$$
$$-e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]\Big(1+3\beta(\alpha-5)+3\beta^2(\alpha-5)(\alpha-2)$$
$$+\beta^3(\alpha-3)(\alpha-2)(\alpha-1)+6e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]$$
$$\times\Big(2\big(1+\beta(-5+2\alpha+(2-3\alpha+\alpha^2)\beta)\big)+3e^{\frac{1}{\beta}}$$
$$\times E[\alpha,\frac{1}{\beta}]\big(2+2\beta(\alpha-1)+e^{\frac{1}{\beta}}E[\alpha,\frac{1}{\beta}]\big)\Big)\Big)\Big\} =: \frac{(a-b)^4 s_3}{6\beta^4}$$

and

$$Skewness(X) = \frac{-s_2}{2\sqrt{s_1}}, \quad Kurtosis(X) = \frac{s_3}{6s_1^2} - 3,$$

respectively.

Note that the skewness and kurtosis measures depend only on α and β.

Table 1. Mean, variance, skewness and kurtosis of $GU(\alpha,\beta,-5,5)$ for various values of α, β.

α	β	E(X)	Var(X)	Skewness(X)	Kurtosis(X)
0.02	0.2	-4.9660	0.0446	10.2131	133.3170
	2	-4.8178	0.7369	6.0870	40.3935
	10	-4.6067	2.2505	4.36755	18.8672
0.5	0.2	-4.2079	0.8622	1.6905	2.9297
	2	-1.5568	7.0084	0.3140	-1.1780
	10	0.9443	9.7743	-0.5957	-1.0360
2	0.2	-2.3945	1.65467	0.3393	-0.4545
	2	2.3073	2.3584	-1.3031	1.7293
	10	4.2015	0.5784	-3.3861	17.4109
10	0.2	1.5189	0.5329	-0.5709	0.4121
	2	4.4769	0.0294	-1.3776	3.7337
	10	4.8903	0.0015	-1.5605	5.0870

From Table 1, it is concluded that for a constant α, the skewness measurement is a decreasing function of β.

4. Percentailes

The pth percentile x_p, is defined by $F(x_p) = p$. From Eq. (2), we have $1 - Q\big(\alpha, \frac{x_p - a}{\beta(b - x_p)}\big) = p$. If $z_{1-p} = \frac{x_p - a}{\beta(b - x_p)}$, then $z_{1-p} = Q^{-1}(\alpha, 1 - p)$, where Q^{-1} is the inverse of regularized incomplete gamma function. Thus $x_p = \frac{a + \beta b z_{1-p}}{1 + \beta z_{1-p}}$.

Example 4.1. If $X \sim GU(2, 4, 0, 5)$ and $p = 0.5$, then $z_{0.5} = Q^{-1}(2, 0.5) = 1.67835$. Therefore the median of the distribution is $x_{0.5} = 4.35178$.

5. Asymptotic properties

Let $X_1, \ldots, X_n$ be a random sample from Eq. (2). Sometimes one would be interested in the asymptotic of extreme values $X_{n:n} = \max(X_1, \ldots, X_n)$ and $X_{1:n} = \min(X_1, \ldots, X_n)$. The limiting distribution of $Y_n = \frac{X_{n:n} - b_n}{a_n}$ is Type I (Exponential type), since $\lim_{n\to\infty} n[1 - F(a_n y + b_n)] = e^{-y}$ in which a_n and b_n are the solution of the system $F(a_n + b_n) = 1 - (ne)^{-1}$ and $F(b_n) = \frac{n-1}{n}$. Hence, it follows from Theorems 7.8.3 and 7.8.5 from [2] that

$$G^{(1)}(y) \approx P(Y_n \le y) = \exp(-e^{-y}) \qquad -\infty < y < \infty$$

and the exact distribution of Y_n is

$$G_n(y) = [F_X(a_n y + b_n)]^n = \big[Q\big(\alpha, \frac{a_n y + b_n - a}{\beta(b - a_n y - b_n)}\big)\big]^n, \quad \frac{a - b_n}{a_n} < y < \frac{b - b_n}{a_n}$$

and the limiting distribution of $W_n = \frac{X_{1:n} + b_n}{an}$ is type III (limiting Type), since $\lim_{n\to 0} \dfrac{1 - F(ky - b_n)}{F(y - b_n)} = k^\alpha$ where $b_n = -\inf\{x|\ F(x) > 0\} = -a$ and $a_n = b_n + s_n$ in which s_n is the solution of $nF(s_n) = 1$. Hence, it follows from Theorem 7.8.6 from [2] that

$$H^{(3)}(w) \approx P(W_n \le w) = 1 - \exp(-w^\alpha), \qquad w > 0$$

and the exact distribution of W_n is

$$H_n(w) = 1 - [1 - F_X(a_n w - b_n)]^n = 1 - \big[1 - Q\big(\alpha, \frac{a_n w}{\beta(b - a_n w - a)}\big)\big]^n, \quad 0 < w < \frac{b - a}{a_n}$$

Example 5.1. Suppose again that $X \sim GU(2, 4, 0, 5)$, and we are interested to obtain the distribution of the maximum and minimum of a random sample of size $n = 100$. It can be easily shown that b_n and a_n is obtained from the solution of the system $1 - Q\big(2, \frac{b_n}{4(5 - b_n)}\big) = \frac{99}{100}$ and $1 - Q\big(2, \frac{a_n + b_n}{4(5 - a_n - b_n)}\big) = 1 - (100e)^{-1}$ which gives $a_n = 0.025748$ and $b_n = 4.81853$.

For minimum of a random sample have $b_n = a = 0$ and s_n is obtained from $1 - Q\big(2, \frac{s_n}{4(5 - s_n)}\big) = \frac{1}{100}$ which gives $s_n = 1.86367$, thus $a_n = 1.86367$.

This is illustrated in Fig. 2, which graphs of $G_n(y)$, $G^{(1)}$, $H_n(w)$ and $H^{(3)}(w)$ for $n = 100$.

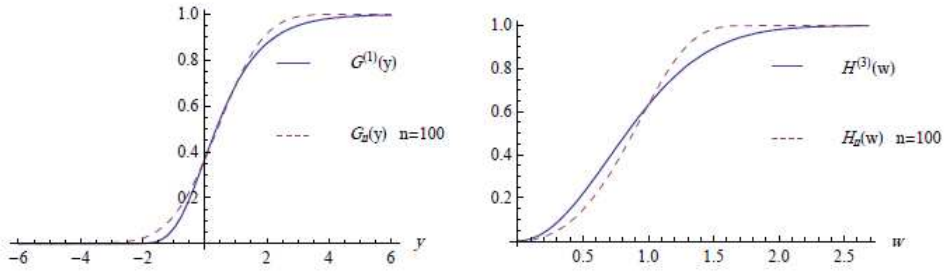

Fig. 2. comparison of cdf $G_n(y)$ with limiting cdf $G^{(3)}(y)$ and cdf $H_n(w)$ with limiting cdf $H^{(1)}(w)$ for $n = 100$.

6. Shannon and Rényi Entropy

An entropy of a random variable X is a measure of variation of the uncertainty. Denote by $\mathcal{H}_{Sh}(f)$ the well-known Shannon entropy introduced by Shannon (1948). It is defined by

$$\mathcal{H}_{sh}(f) = E[-\log f(X)] = -\int_{\mathcal{X}} f(x)\log(f(x))\ dx \tag{5}$$

One of the main extensions of was defined by Rényi (1961). This generalized entropy measure is given by $\mathcal{H}_R(\lambda) = \mathcal{H}_R(\lambda, f) = \frac{\log G(\lambda)}{1-\lambda}$ for $\lambda > 0$ *and* $\lambda \neq 1$, where

$$G(\lambda) = \int_{\mathcal{X}} f(x)^{\lambda}\ dx \tag{6}$$

The additional parameter λ is used to describe complex behavior in probability models and the associated process under study. Rényi entropy $\mathcal{H}_R(\lambda)$ is monotonically decreasing in λ while Shannon entropy Eq. (5) is obtained from Eq. (6) for $\lambda \uparrow 1$.[10]

Theorem 6.1. *The Shannon and Rényi Entropy of Gamma-Uniform distribution are equal with the Shannon and Rényi Entropy of Gamma distribution.*

Proof. By taking $w = \frac{x-a}{b-x}$ in Eq. (5) and Eq. (6) have;

$$\mathcal{H}_{sh}(f) = -\int_0^{\infty} \frac{e^{-\frac{w}{\beta}} w^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} \log\left(\frac{e^{-\frac{w}{\beta}} w^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}}\right) dw = \alpha + \log\big(\beta\Gamma(\alpha)\big)(\alpha-1)\psi(\alpha)$$

where $\psi(z) = \frac{d \ln \Gamma(z)}{dz} = \frac{\Gamma'(z)}{\Gamma(z)} =$ is Poly Gamma function. Rényi entropy is

$$\begin{aligned}\mathcal{H}_R(\lambda) &= \frac{1}{1-\lambda} \log \big(\int_0^\infty \big(\frac{e^{-\frac{w}{\beta}} w^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \big)^\lambda dw \big) \\ &= \frac{1}{1-\lambda} \Big[\log \Gamma(\lambda\alpha - \lambda + 1) - \lambda \log \Gamma(\alpha) - (\lambda+1) \log \beta \\ &\quad -(\lambda\alpha - \lambda + 1) \log \lambda \Big], \quad \text{if } r < \tfrac{1}{1-\alpha} \text{ or } \alpha > 1. \end{aligned}$$ □

7. Simulation

For simulation of the distribution, note from Eq. (2) that if w is a random number from a $G(\alpha, \beta)$, then $\frac{a+bw}{1+w}$ will follow the pdf of Eq. (2).

8. Conclusions

In this paper, a new flexible class of distributions is considered, then for a special case of this class, some moments, Shannon and Rényi entropy are derived. Finally, the asymptotic distribution of the extreme order statistics and simulation issues are discussed.

References

1. A. Akinsete, F. Famoye and C. Lee. The Beta-Pareto distribution. *Statistics* 2008;**42**:547-563.
2. L.J. Bain and M. Engelhardt.(1991). Introduction to probability and Mathematical Statistics, 2nd Edition, Duxbury press.
3. Wagner Barreto-Souza, Alessandro H.S. Santos and Gauss M. Cordeirob. The Beta generalized exponential. *Statistical Computation and Simulation distribution* 2009;**0**, In Press.
4. N. Eugene, C. Lee, F. Famoye. The Beta-Normal distribution and its applications. *Communication in Statistics: Theory and Methods* 2002;**31**:497-512.
5. C. Lee, F. Famoye, O. Olumolade. The Beta-Weibull distribution: Some properties and applications to censored data. *Journal of Modern Applied Statistical Methods* 2007:**6**:173-186.
6. S. Nadarajah, A.K. Gupta. The Beta Fréchet distribution. *Far East Journal of Theorical Statistics* 2004;**15**:15-24.
7. S. Nadarajah, S. Kotz. The Beta Gumbel distribution. *Mathematical Problems in Engineering* 2004;**10**:323-332.
8. S. Nadarajah, S. Kotz. The Beta exponential distribution. *Reliability Engineering and System Safety* 2006;**91**:689-697
9. Rodrigo R. Pescim, Clarice G.B. Demtrio, Gauss M. Cordeiro, Edwin M.M. Ortega, Mariana R. Urbano. The Beta generalized half-Normal distribution. *Computational Statistics and Data Analysis* 2009;**54**:945-957.
10. K. Zografos, S. Nadarajah. Expressions for Rényi and Shannon entropies for multivariate distributions. Statistics and Probability Letters 2005;**71**:71-84.

Applications of Bayesian Point Null Hypothesis Testing Via the Posterior Likelihood Ratio

E. Torkashvand[1], M. F. Rohani[1] and M. Ganjali[2]

M. Ganjali

[1] *Department of Statistics, Shahid Beheshti University*

[2] *Department of Statistics, Shahid Beheshti University*

This paper introduces the application of the posterior likelihood ratio for hypothesis testing in 2×2 contingency tables. The analysis has been done by using of different prior distributions and the log of Odds Ratio. It also evaluates another application of the posterior likelihood ratio for hypothesis testing in nonnested models, such as Geometric against Poisson and Gamma against Weibull. To make inferences, this paper uses noninformative prior distributions for all parameters independently from previous trials. Because the posterior distributions are not well-known, and as a result, they cannot be simulated easily, Acceptance-Rejection and Metropolis-Hastings algorithms and also Gibbs sampler are used to obtaining random draws from exactly the posterior distributions.

Keywords: Posterior Distribution; Likelihood Ratio; Fisher Exact Test; Nonnested Models; Metropolis-Hastings Algorithm; Acceptance-Rejection Algorithm; Gibbs Sampler.

1. Introduction

In the point null hypothesis testing the results of Bayes factor and P-value approach are not consistent, so we consider a new approach in hypothesis testing, the posterior likelihood ratio. The results of this approach are consistent with frequentist P-value conclusions. Section 2 of the paper illustrates the general approach of the posterior likelihood ratio in the presence of nuisance parameters, and analyzes it more. As a new practical use Section 3 introduces the application of this new approach in Fisher's Exact test. We use the log of Odds Ratio, $\delta = \log(\psi)$, to examine the relation between rows in 2×2 contingency tables. Moreover, we make use of Fisher's famous example with informative and noninformative prior distributions to represent how this approach works.

The other new application of the posterior likelihood ratio in nonnested models such as Geometric against Poisson and Gamma against Weibell is evaluated in Section 4. . We use data which are provided by R software to deal with an extensive

domain of data. Finally, we illustrate the conclusions in Section 5.

2. The Posterior Likelihood Ratio

Let θ be an interesting parameter and ϕ be a nuisance parameter. So, for the given data, $\underline{y}$, the likelihood of these parameters is:

$$L(\theta, \phi) = f(\underline{y} \mid \theta, \phi).$$

The prior distribution for these parameters represents our uncertainty about these parameters. Consider the point null hypothesis, H_0, which specifies the value of θ_0. An alternative hypothesis, H_1, specifies either that θ is completely unspecified, $\theta \neq \theta_0$, or that θ has a different specified value θ_1. In either case ϕ is unspecified. The likelihood ratio between the null and alternative hypothesis is defined in the following way:

$$LR = LR(\theta, \phi) = \frac{L(\theta_0, \phi)}{L(\theta, \phi)},$$

where for simplicity the dependency of LR on θ_0 and $\underline{y}$ is ignored.

Suppose $\pi(\theta, \phi \mid \underline{y})$ be the joint posterior distribution for θ and ϕ. As LR is a function of ϕ and θ, we can calculate the posterior distribution of LR, especially the posterior probability $P(LR < k \mid \underline{y})$ for any specific k, such as 0.1 or 0.01.

Although this approach is different from classical testing, the form of the posterior likelihood ratio, (γ, k) where $\gamma = P(LR < k)$, for hypothesis testing has the same results as the significant test (Dempster, 1974-1997). But the posterior likelihood ratio is more conservative than the significant test. When a researcher knows $LR \geq 1$, he cannot accept the alternative hypothesis. For example, when an observed data is in the boundary points of the rejection region,a frequentist rejects the null hypothesis, but in the posterior likelihood ratio approach, we need stronger evidence, which is provided by the posterior probability $LR \geq k$ to reject the null hypothesis.

Dempster(1974-1997) and Aitkin(1997) for uniform priors and normal likelihoods showed that:

$$P(LR < 1 \mid \underline{y}) = 1 - P - value. \tag{1}$$

In classical framework, when P-value is smaller than the significant level , we evaluate the existing evidence fine. For example, when:

$$0 \leq P - value \leq 0.05$$

according to existing evidence, we reject the null hypothesis at 0.05 significant level. The aim of introducing the posterior likelihood ratio was the consistency between the results of this approach and the results of classical approach (Aitkin et.al., 2005). Therefore, in relation to (1), we obtain following limits for $P[LR < 1 \mid \underline{y}]$, corresponding to 0.05 significant level:

$$0.95 \leq P[LR < 1 \mid \underline{y}] < 1.$$

On the other hand, we emphasized that the posterior likelihood ratio is more conservative than classical approach. One of the reasons which causes to think about the conservative structure of this approach is that we can calculate posterior probabilities $P[LR < k \mid \underline{y}]$ for all $0 < k < 1$ and care about them. Actually, this approach assesses the existing evidence to reject the null hypothesis insufficient, when the probability of $\mathrm{P}[LR < 0.1 \mid \underline{y}]$ is small.

Now, we evaluate the mathematical structure of LR more precisely. For the point null hypothesis, we have:

$$LR(\theta) = LR(\theta) = \frac{L_{H_0}(\theta)}{L_{H_1}(\theta)} = \frac{L(\theta_0)}{L_{H_1}(\theta)}.$$

For the compound alternative hypothesis, the minimum value of LR is obtained when $L_{H_1}(\theta)$ gets its maximum value, i.e. when $\theta = \theta_{ML}$. Therefore:

$$\theta_{ML} \in \{\theta; LR(\theta) < k\},$$

and this set also contains the most probable points of $\Theta - \{\theta_0\}$ in respect to the observed data. As a result, when $P[LR < k \mid \underline{y}]$ is a considerable probability, existing evidence convinces us to reject the null hypothesis. Although the minimum level of the posterior probability in order to reject or accept the null hypothesis is chosen by the researcher according to the importance of the null hypothesis, one may count probabilities such as 0.25 or more as a large probability(Aitkin et. al. 2005)

Another important point which is introduced is the limits of k. As this paper reminds, Fisher(1958) made use of the likelihood to find confidence intervals. Fisher's main goal was the usage of these functions to obtain confidence intervals when a pivotal quantity was not available. Actually, he used the MLE to specify the inner points of the interval, around that and the excluded points of the parameter space with likelihoods less than $\frac{1}{15}$ of the MLE. In the posterior likelihood ratio, similarly, Aitkin(1997) and Aitkin et.al.(2005) specified values of the parameter space with $LR < 0.1$ as parameters which are suspicious to be accepted as the true value of the population parameter.

3. Application of the posterior likelihood ratio in 2×2 contingency tables

When the sample size is large, classical framework in hypothesis testing has a great power. But sometimes in practice, we have to work with the small sample size. In this case, classical approaches are too conservative.

In this section, 2×2 contingency tables are evaluated by Dempster's approach with different prior distributions. We use Metropolis-Hastings and Acceptance-Rejection algorithm to obtain random samples from the posterior distributions.

Consider a hypothesis testing which appraises the hypothesis of no association between rows and columns in a 2×2 contingency table versus the hypothesis of positive association between them. In general, conditional distribution of N_{11}

statistics given their marginal values is the noncentral Hypergeometric as follows (Fisher(1935) and Agresti(2002)):

$$f(n_{11} \mid n, n_{1+}, n_{+1}; \theta) = \frac{\binom{n_{1+}}{n_{11}}\binom{n_{2+}}{n_{+1}-n_{11}}\theta^{n_{11}}}{\sum_u \binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1}-u}\theta^{u}},$$

where

$$u \in \{\max\{0, n_{+1} + n_{1+} - n\}, ..., \min\{n_{1+}, n_{+1}\}\},$$

and $\theta > 0$ is Odds Ratio. If we define δ as the log of Odds Ratio($\ln\theta$),therefore it is equivalent to zero when the variables are not associated. In this case, the noncentral Hypergeometric distribution converts to the central Hypergeometric distribution. Under the null hypothesis δ is also symmetric around zero. This reparameterization helps us to analyze the use of the posterior likelihood ratio under different prior distributions on δ such as $N(\mu, \sigma^2)$ and the noninformative prior distribution and also conjugate prior distributions of Exponential family like:

$$K(\eta, \tau)\frac{\exp(\delta\eta)}{[\sum_u \binom{n_{1+}}{u}\binom{n_{2+}}{n_{+1}-u}\exp(\delta u)]^{\tau}} \tag{2}$$

where η and τ are hyperparameters, K is the normalizing constant, and

$$u \in \{\max\{0, n_{+1} + n_{1+} - n\}, ..., \min\{n_{1+}, n_{+1}\}\}.$$

3.1. *Applied Example*

Consider Fisher's famous example. We want to examine the null hypothesis, $H_0 : \theta = 1$ versus the alternative hypothesis, $H_1 : \theta > 1$.

At first, we assess the conjugate prior distributions (2) which define probability density function on $\delta \in [0, \infty)$.

According Berger and Robert (1990), we know that using hirachical Bayesian models with flat prior distributions on hyperparameters provides robust Bayesian estimators. Therefore, we apply the hirechical Bayesian approach with flat priors on η and τ in order to solve the problem. We use uniform prior distributions on η and τ to obtain robust Bayesian estimators in this problem:

$$\eta \sim Uniform(-100, -1),$$
$$\tau \sim Uniform(1, 6).$$

Because we cannot generate random draws from the conjugate prior distributions directly, we use Acceptance-Rejection algorithm to obtain random draws from them. At first, we need to find an envelope . Our choice is an exponential distribution with mean one. After that, we should find a constant c, such that for all pairs of (η, τ) the envelope , $c\exp(-\delta)$, as a coverage for all prior distributions. By simplicity we can find the following condition for constant c:

$$c \geq 106 \times K(-100, 6).$$

Now, we generate samples from the hirechical prior distribution. We apply the following Gibbs sampler to obtain random variables from the hirechical prior distribution.

1. Generate $\eta \sim Uniform(-100, -1)$ and independently $\tau \sim Uniform(1, 6)$,
2. Simulate $K(\eta, \tau)$ from Importance sampling method where the proposed distribution is Gamma(1,1),
3. Generate δ from (2) with above parameter (η, τ) by the Acceptance-Rejection algorithm.

The next steps show how Acceptance-Rejection algorithm is used to obtaining random draws:

- Draw a sample $y \sim \exp(-\delta)$,
- Draw a sample $t \sim Uniform(0, c\exp(-y))$,
- If $K(\eta, \tau)\frac{\exp(\eta\delta)}{(1+4\exp(\delta)+\exp(2\delta))^{\tau}} \leq t$, accept y as a random draw from the conjugate prior distribution with known parameters (η, τ) and come back to the first step. Otherwise, reject y and come back to the first step.

We input δ's which are samples from Gibbs sampler into Metropolis-Hastings algorithm in order to get random draws from the posterior distribution. The simulated probabilities are:

$$P(LR < 0.1) \approx 0, P(LR < 0.167) \approx 0.005, P(LR < 0.375) \approx 0.5.$$

Truncated $N(\mu, \sigma^2)$ distribution at zero is evaluated here as another prior distribution(this distribution assigns probabilities on $\delta \in [0, \infty)$ as same as the conjugate prior distribution). The parameters of the truncated $N(\mu, \sigma^2)$ are computed according to the following formula for Fisher's example:

$$\log(\psi) \sim N(\log(\hat{\psi}), ASE(\log(\hat{\psi}))),$$

where

$$ASE(\log(\hat{\psi})) = \sqrt{\frac{1}{n_{11}+\frac{1}{2}} + \frac{1}{n_{12}+\frac{1}{2}} + \frac{1}{n_{21}+\frac{1}{2}} + \frac{1}{n_{22}+\frac{1}{2}}}.$$

Thus, we have truncated $N(3.219, 4.8)$ at zero. As the null hypothesis is accepted with two previous prior distributions, we decide to assess the importance of information which is provided by prior distributions. To gain this purpose, we put noninformative distribution $\pi(\delta) = 1$ on $-\infty \leq \delta \leq \infty$. As same as the previous prior distributions, noninformative prior distribution confirms the null hypothesis.

4. The nonnested hypothesis testing via the posterior likelihood ratio

Generally, our aim in statistics is to make inferences about some quantities of interest, Δ, on the basis of some observed data $\underline{X}$ and $\underline{Y}$(the quantity Δ may

be a regression parameter, a forecast of some future observation, or some other quantity of interest). In classical statistical theory, the data are only informative about Δ when they are interpreted in light of definite statistical model, M, which specifies a relevant population, a set of relevant variables, the functional form of the relationships among these variables, and the nature of relevant stochastic influences. All of our inferences about Δ depend not only upon the data $\underline{X}$ and $\underline{Y}$, but also assumptions embodied in the model M. When there is some doubt about the model of stochastic process, it seems necessary to test models in order to choose the best one which shows the stochastic process well. Sometimes these models are from different families. In this case, we want to test nonnested models.

In this section, nonnested models are evaluated via the posterior likelihood ratio by using noninformative prior distributions on all parameters.

4.1. *Applications of the posterior likelihood ratio in nonnested models*

4.1.1. *Geometric versus Poisson*

Suppose that we want to evaluate the null hypothesis of Geometric distribution with unknown parameter θ versus the alternative hypothesis of Poisson distribution with unknown parameter λ. In this case the likelihood ratio is:

$$LR(\theta, \lambda) = \frac{\theta^n (1-\theta)^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n \frac{1}{x_i!} e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}. \tag{3}$$

For calculating the posterior likelihood ratio probabilities, it is enough to generate θ and λ values from their own posterior distributions and to substitute them in (3). We use Metropolis-Hastings algorithm to generate from these posterior distributions.

When we simulate data from Geometric distribution, proposal approach works as well as possible. Even by interchanging the null and the alternative hypotheses, there is not any contradictions to accept Geometric distribution. An interesting point occurs when we evaluate the generated data from Poisson distribution. When the true value of $\lambda > 1$, the posterior likelihood ratio approach determines the true distribution correctly. But when $\lambda \leq 1$, we face a problem. In this case, our approach chooses Geometric distribution or both of Geometric and Poisson distributions while we interchange the null and the alternative hypotheses.

Because the prior distributions are improper, we can neither compute Bayes Factor nor inference based on it.

4.1.2. *Gamma versus Weibull*

Another example of nonnested models is Gamma distribution where α is the interesting parameter and β is the nuisance parameter versus Weibull distribution where δ is the interesting parameter and λ is the nuisance parameter. We get the

likelihood ratio as:

$$LR(\alpha,\delta)=\frac{\frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}(\prod_{i=1}^n x_i)^{(\alpha-1)}\exp(-\beta\sum_{i=1}^n x_i)}{\delta^n\lambda^n(\prod_{i=1}^n x_i)^{(\delta-1)}\exp(-\lambda\sum_{i=1}^n x_i^\delta)}$$

As usual, we put flat prior distributions on $\alpha>0,\beta>0,\delta>0$, and $\lambda>0$. We find out that the posterior distributions of this example are not well known. Thus, we need to make use of Gibbs sampler to generate random variables.

The resulted outcomes are not consistent with our data. This means that although data which is provided by Gamma or Weibull distributions, respectively is tested and accepted by the null hypothesis of Gamma and Weibull distributions, by interchanging the null and the alternative hypotheses , it is also accepted as the null hypothesis. Therefore, the application of the posterior likelihood ratio in this problem is as same as Cox's test. Both of them cannot determine the true distribution when we interchange the null and the alternative hypotheses. Thus, whenever we use the posterior likelihood ratio approach in nonnested models hypothesis testing, we should choose the null hypotheses carefully.

We cannot compute Bayes Factor as the previous example of the nonnested models because of improper prior distributions.

5. Conclusions

As we have considered, the posterior likelihood ratio is a new approach which needs to make a lot of efforts in order to improve it. In this article, we analyze the posterior likelihood ratio more mathematically.

On the other hand, Aitkin et. al.(2005) emphasized when Classical approach and Bayes Factor are in contradiction, the posterior likelihood ratio leads to outcomes similar to Classical approach. As we show in this paper, the null hypothesis of no association is accepted by the posterior likelihood ratio approach with different prior distributions, which is the same result as P-value in Classical approach.

One of the most important advantages of this approach is the ability of making a Bayesian test when we cannot calculate Bayes Factor when the prior distributions are improper. As we see, Bayes Factor cannot be calculated in both examples of nonnested models. But by using the posterior likelihood approach, we made Bayesian inferences. Outcomes for the test of nonnested models Geometric against Poisson choose the correct distribution of data when the true distribution is Geometric, but when we obtain data from Poisson distribution, this approach selects the true model under some conditions. In addition, the proposal approach is not capable enough to choose the true model when we test Gamma distribution versus Weibull distribution and interchange the null the alternative hypotheses.

References

1. A. Agresti, *Categorical Data Analysis*, (J. Willey, New York., 2002).

2. M. Aitkin, *Statist and Computing* **7,** The calibration of P-values, posterior Bayes Factors and the AIC from the posterior distribution of the likelihood(with discussion). (1997).
3. M. Aitkkin, R. J. Boys and T. Chadwick, *Statistics and Computing* **15,** Bayesian point null hypothesis testing via the posterior likelihood ratio. (2005).
4. J. O. Berger and C. Robert, *Ann. Statist.* **18,** Subjective hirechical Bayes estimation of a multivariate normal mean: On the frequentist interface. (1990).
5. A. P. Dempster, In: Barndorff-Nielsen O., Blaesild P. and Sihon G. (Eds.), *Proc. Conf. Foundational Questions in Statistical Inference pp.*, The direct use of likelihood in significance testing. (1974).
6. A. P. Dempster, *Statist. and Computing* **7,** The direct use of likelihood in significance testing. (1997).
7. R. A. Fisher, *J. R. Statist. Soc. A* **98,** The logic of inductive inference. (1935).
8. R. A. Fisher, *Statistical methods and scientific inference*, 2nd ed. (Oliver and Boyd, Edinburgh, 1958).

Sample Entropy Of Continuous Distributions

F. Yousefzadeh and N. R. Arghami

Department of Statistics, Birjand University

Department of Statistics, Ferdowsi University of Mashhad

This paper improves the existing nonparametric estimators, of the entropy of a continuous distribution, using a proposed nonparametric cdf estimator. We prove the consistency of the proposed estimator and show, by simulation, that our estimator has less bias and less root mean squared error (RMSE) than the leading estimators proposed by Ebrahimi et al(1994).

Keywords: Entropy estimator, cdf estimator, Kernel density.

1. Introduction

Suppose X is distributed according to a continuous distribution with cumulative distribution function(cdf) F, having support $I = [a_0, b_0]$ where $a_0 \geq -\infty$, $b_0 \leq \infty$. The entropy of F is

$$H(X) = -\int f(x)\ln(f(x))dx,$$

where f is the probability density function (pdf) of X.
The entropy estimator of F has been discussed by many authors including Ahmad and Lin (1976),Vasicek (1976), Mack (1988), Joe(1989), Makkadem(1989) and Ebrahimi et al.(1992).
Ebrahimi et al.(1994) proposed two estimators of the entropy in the following forms:

$$H_c = \frac{1}{n}\sum_{i=1}^{n}\ln(\frac{x_{i+m} - x_{i-m}}{c_i m/n}), \tag{1}$$

where

$$c_i = \begin{cases} 1 + \frac{i-1}{m} & 1 \leq i \leq m, \\ 2 & m+1 \leq i \leq n-m, \\ 1 + \frac{n-i}{m} & n-m+1 \leq i \leq n. \end{cases}$$

Here $x_1, ..., x_n$ are ordered values of random observation of X and m is a positive integer less than or equal to $n/2$, which is called window size. We take $x_{i-m} = x_1$

for $i-m<1$ and $x_{i+m}=x_n$ for $i+m>n$.
The second estimator of Ebrahimi et al. is

$$H_d=\frac{1}{n}\sum_{i=1}^{n}\ln(\frac{z_{i+m}-z_{i-m}}{d_i m/n}), \tag{2}$$

where

$$d_i=\begin{cases}1+\frac{i+1}{m}-\frac{i}{m^2} & 1\le i\le m,\\ 2 & m+1\le i\le n-m,\\ 1+\frac{n-i}{m+1} & n-m+1\le i\le n,\end{cases}$$

and z_i's are

$$\begin{aligned}&z_{i-m}=a+\tfrac{i-1}{m}(x_1-a)=x_1-\tfrac{m-i+1}{m}(x_1-a), && 1\le i\le m,\\ &z_i=x_i, && m+1\le i\le n-m-1,\\ &z_{i+m}=b-\tfrac{n-i}{m}(b-x_n)=x_n+\tfrac{m+i-n}{m}(b-x_n), && n-m\le i\le n,\end{aligned}$$

where a and b are equal to a_0 and b_0, respectively, except when $a=-\infty(b=\infty)$, in which case $a(b)$ is taken to be $\bar{x}-ks$ $(\bar{x}+ks)$, where

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i, \qquad s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

where k is an integer(usually between 3 and 5), which minimizes the simulated RMSE.
In this paper we use a new estimator of F. In Section 2 we introduce a new cdf estimator which can be regarded a constrained MLE of F. In Section 3, using a new cdf estimator, we introduce an estimator of the entropy of F and show that it is consistent and that the bias and mean squared error of this estimator is scale invariant. Section 4 contains the results of a simulation study which shows that our entropy estimators compare favorably with those of Ebrahimi et al.(1994)'s.

2. A New cdf Estimator

Let $x_1,...x_n$, F and f be as in Section 1. Let

$$P(x_{i-1}<X<x_i)=w_i, \qquad i=2,...,n,$$

approximating $P(X<x_1)$ and $P(X>x_n)$ by $\frac{1}{n+1}$.
We impose the condition

$$\sum_{i=2}^{n}w_i=\frac{n-1}{n+1}. \tag{3}$$

Here w_i is taken to be the area, under the plot of f, between x_{i-1} and x_i. Then we can approximate w_i by

$$w_i\approx\frac{f_{i-1}+f_i}{2}d_i, \tag{4}$$

where $d_i = x_i - x_{i-1}$, $f_i = f(x_i)$ and we define $d_1 = d_{n+1} = 0$. We maximize the likelihood function subject to (4). For this purpose we write the Lagrangian as

$$L = \sum_{i=1}^{n} f_i - 2\lambda[\sum_{i=2}^{n} w_i - \frac{n-1}{n+1}],$$

which leads us to solving the equations

$$\frac{\partial L}{\partial f_i} = 0, \qquad \forall i = 1, ..., n,$$

which yeild

$$f_i = \frac{1}{\lambda(d_i + d_{i+1})},$$

$$\lambda = \frac{n+1}{2(n-1)} \sum_{i=2}^{n} (\frac{1}{d_i + d_{i+1}} + \frac{1}{d_{i-1} + d_i}) d_i = \frac{n(n+1)}{2(n-1)},$$

and

$$w_i = \frac{n-1}{n(n+1)} (\frac{1}{d_{i-1} + d_i} + \frac{1}{d_i + d_{i+1}}) d_i, \qquad i = 1, ..., n,$$

which leads to

$$\hat{F}_n(x_i) = \sum_{j=1}^{i} w_j = \frac{n-1}{n(n+1)} (i + \frac{1}{n-1} + \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}}),$$

as estimator of $F(x_i)$'s, $i = 2, ..., n-1$. Considering (3) we have $F(x_1) = \frac{1}{n+1}$, $F(x_n) = \frac{n}{n+1}$.
We use the linear interpolation for $x < x_1$, and for $x > x_n$ as follow

$$\hat{F}_n(x) = \begin{cases} \frac{1}{n+1} \frac{x-a}{x_1 - a} & a \le x \le x_1, \\ \\ \frac{1}{n+1} (n + \frac{x - x_n}{b - x_n}) & x_n \le x \le b, \end{cases} \tag{5}$$

where a and b are as in Section 1.

Theorem 2.1. *Let $x_1, ..., x_n$ be an ordered random sample from a distribution with cdf F. Then $\hat{F}_n(x)$ is a consistent estimator of $F(x)$.*

Proof. We can write $\hat{F}_n(x)$ as

$$\hat{F}_n(x_i) = \frac{n-1}{n(n+1)} (i + \frac{1}{n-1} + \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}}), \quad , \quad i = 1, ..., n.$$

So we have

$$\hat{F}_n(x_i) \le \frac{n-1}{n(n+1)} (i + \frac{1}{n-1} + 1) \le \frac{n-1}{n(n+1)} (i+2), \quad , \quad i = 1, ..., n.$$

Then

$$\hat{F}_n(x) \leq \frac{n-1}{n+1}(F_n(x) + \frac{2}{n}), \qquad x_1 \leq x \leq x_n,$$

where $F_n(x)$ is empirical distribution function. Similarly

$$\hat{F}_n(x) \geq \frac{n-1}{n+1}(F_n(x)). \qquad x_1 \leq x \leq x_n.$$

So we conclude that

$$E(\hat{F}_n(x)) \longrightarrow F(x), \qquad Var(\hat{F}_n(x)) \longrightarrow 0,$$

and thus $\hat{F}_n(x)$ is consistent. □

3. Entropy Estimator

Using the above estimator of F, analogous to (1) and (2), we propose two estimators for the entropy $H(X)$ of an unknown continuous distribution F, by

$$H_a(m,n) = \frac{1}{n}\sum_{i=1}^{n} \ln \frac{x_{i+m} - x_{i-m}}{\hat{F}_n(x_{i+m}) - \hat{F}_n(x_{i-m})} \tag{6}$$

and

$$H_b(m,n) = \frac{1}{n}\sum_{i=1}^{n} \ln \frac{z_{i+m} - z_{i-m}}{\hat{F}_n(z_{i+m}) - \hat{F}_n(z_{i-m})}, \tag{7}$$

where x_i andz_i are as in Section 1.
The following theorem is used in the proof of Theorem 3.2.

Theorem 3.1 (Vasicek, 1976). *Let $x_1, ..., x_n$ be an ordered sample from a distribution F with a density f and a finite variance. Then*

$$H'_{mn} \xrightarrow{p} H(X) \ \ as \ \ n, m \to \infty, \ \ such \ that \ \ m/n \to 0,$$

where $H'_{mn} = H_{mn} - E(U_{mn})$,where

$$\begin{aligned} H_{mn} &= \tfrac{1}{n}\sum_{i=1}^{n} \log(\tfrac{x_{i+m}-x_{i-m}}{2m/n}) \\ &= -\tfrac{1}{n}\sum_{i=1}^{n} \log f(x_i) + V_{mn} + U_{mn}, \end{aligned}$$

where

$$V_{mn} = \frac{1}{n}\sum_{i=1}^{n} \log(\frac{F(x_{i+m}) - F(x_{i-m})}{f(x_i)(x_{i+m} - x_{i-m})},$$

$$U_{mn} = \frac{1}{n}\sum_{i=1}^{n} \log(\frac{F(x_{i+m}) - F(x_{i-m})}{2m/n}).$$

The consistency of H_a and H_b are proved by the following theorem.

Theorem 3.2. *Let $x_1, ..., x_n$ be an ordered random sample from a continuous distribution F. then as $n, m \to \infty$, such that $\frac{m}{n} \to 0$,*

- *$H_a(m,n) \to H(X)$ in probability*
- *$H_b(m,n) \to H(X)$ in probability.*

Proof.

- In view of Theorems 2.1 and 3.1, the proof is easy and is thus omitted.
- We can write

$$\begin{aligned} H_b(m,n) &= \tfrac{1}{n}\sum_{i=1}^{n} \ln \tfrac{z_{i+m}-z_{i-m}}{\hat{F}_n(z_{i+m})-\hat{F}_n(z_{i-m})} \\ &= \tfrac{1}{n}\sum_{i=1}^{m} \ln \tfrac{x_{i+m}-z_{i-m}}{\hat{F}_n(x_{i+m})-\hat{F}_n(z_{i-m})} + \tfrac{1}{n}\sum_{i=m+1}^{n-m-1} \ln \tfrac{x_{i+m}-x_{i-m}}{\hat{F}_n(x_{i+m})-\hat{F}_n(x_{i-m})} \\ &+ \tfrac{1}{n}\sum_{i=n-m}^{n} \ln \tfrac{z_{i+m}-x_{i-m}}{\hat{F}_n(z_{i+m})-\hat{F}_n(x_{i-m})}, \end{aligned}$$

also

$$\begin{aligned} H_a(m,n) &= \frac{1}{n}\sum_{i=1}^{n} \ln \frac{x_{i+m}-x_{i-m}}{\hat{F}_n(x_{i+m})-\hat{F}_n(x_{i-m})} \\ &= \frac{1}{n}\sum_{i=1}^{m} \ln \frac{x_{i+m}-x_1}{\hat{F}_n(x_{i+m})-\hat{F}_n(x_1)} + \frac{1}{n}\sum_{i=m+1}^{n-m-1} \ln \frac{x_{i+m}-x_{i-m}}{\hat{F}_n(x_{i+m})-\hat{F}_n(x_{i-m})} \\ &+ \frac{1}{n}\sum_{i=n-m}^{n} \ln \frac{x_n-x_{i-m}}{\hat{F}_n(x_n)-\hat{F}_n(x_{i-m})}. \end{aligned}$$

Thus

$$\begin{aligned} &H_b(m,n) - H_a(m,n) = \\ &\frac{1}{n}\sum_{i=1}^{m} \ln \frac{x_{i+m}-z_{i-m}}{\hat{F}_n(x_{i+m})-\hat{F}_n(z_{i-m})} - \frac{1}{n}\sum_{i=1}^{m} \ln \frac{x_{i+m}-x_1}{\hat{F}_n(x_{i+m})-\hat{F}_n(x_1)} \\ &+ \frac{1}{n}\sum_{i=n-m}^{n} \ln \frac{z_{i+m}-x_{i-m}}{\hat{F}_n(z_{i+m})-\hat{F}_n(x_{i-m})} - \frac{1}{n}\sum_{i=n-m}^{n} \ln \frac{x_n-x_{i-m}}{\hat{F}_n(x_n)-\hat{F}_n(x_{i-m})} \\ &= \frac{1}{n}\sum_{i=1}^{m} \ln\left(\frac{x_{i+m}-z_{i-m}}{x_{i+m}-x_1}\,\frac{\hat{F}_n(x_{i+m})-\hat{F}_n(x_1)}{\hat{F}_n(x_{i+m})-\hat{F}_n(z_{i-m})}\right) \end{aligned}$$

$$+\frac{1}{n}\sum_{i=n-m}^{n}\ln\left(\frac{z_{i+m}-x_{i-m}}{x_n-x_{i-m}}\frac{\hat{F}_n(x_n)-\hat{F}_n(x_{i-m})}{\hat{F}_n(z_{i+m})-\hat{F}_n(x_{i-m})}\right)$$
$$\leq \frac{1}{n}\sum_{i=1}^{m}\ln\frac{x_{i+m}-z_{i-m}}{x_{i+m}-x_1}+\frac{1}{n}\sum_{i=n-m}^{n}\ln\frac{z_{i+m}-x_{i-m}}{x_n-x_{i-m}}.$$

Ebrahimi et al.(1994) show that the right hand side of the above inequality goes to zero in probability. Also we can write

$$H_b(m,n)-H_a(m,n)$$
$$\geq \frac{1}{n}\sum_{i=1}^{m}\ln\frac{\hat{F}_n(x_{i+m})-\hat{F}_n(x_1)}{\hat{F}_n(x_{i+m})-\hat{F}_n(z_{i-m})}+\frac{1}{n}\sum_{i=n-m}^{n}\ln\frac{\hat{F}_n(x_n)-\hat{F}_n(x_{i-m})}{\hat{F}_n(z_{i+m})-\hat{F}_n(x_{i-m})}$$
$$= I+II,$$

where $I=\frac{1}{n}\sum_{i=1}^{m}A_i$ and $II=\frac{1}{n}\sum_{i=n-m}^{n}B_i$, where

$$A_i=\frac{\hat{F}_n(x_{i+m})-\hat{F}_n(x_1)}{\hat{F}_n(x_{i+m})-\hat{F}_n(z_{i-m})} \quad and \quad B_i=\frac{\hat{F}_n(x_n)-\hat{F}_n(x_{i-m})}{\hat{F}_n(z_{i+m})-\hat{F}_n(x_{i-m})}.$$

It is easy to see that

$$A_i\geq\frac{\frac{n-1}{n}[i+m+\frac{1}{n-1}]-1}{\frac{n-1}{n}[i+m+1+\frac{1}{n-1}]-\frac{i-1}{m}}\geq\frac{m}{2m+\frac{n}{n-1}}$$

and

$$B_i\geq\frac{n-\frac{n-1}{n}[i-m+\frac{1}{n-1}+1]}{n+\frac{m+i-n}{m}-\frac{n-1}{n}[i-m]}\geq\frac{m}{2m+2\frac{n}{n-1}}.$$

So, as $n,m\to\infty$ $s.t.$ $\frac{m}{n}\to 0$

$$I\geq\frac{m}{n}\ln\frac{m}{2m+\frac{n}{n-1}}\to 0 \quad and \quad II\geq\frac{m}{n}\ln\frac{m}{2m+2\frac{n}{n-1}}\to 0.$$

Thus the result follows. □

The next theorem shows that the scale of the random variable X has no effect on the accuracy of $H_a(m,n)$ and $H_b(m,n)$ in estimating $H(X)$.

Theorem 3.3. *Let $H(X)$ and $H(W)$ denote entropies of X and W, respectively, and let $X_1,X_2,...,X_n$ be a an ordered random sample from a distributed with cdf F and $W_i=kX_i$, $i=1,...,n$, where $k>0$. Then the following relations hold.*

1. $E(H_a^W(m,n)) = \ln k + E(H_a^X(m,n))$,
2. $Var(H_a^W(m,n)) = Var(H_a^X(m,n))$,
3. $MSE(H_a^W(m,n)) = MSE(H_a^X(m,n))$,

where the superscripts X and W refer to the corresponding distribution.

Proof. It is easy to see that

$$\begin{aligned}\hat{F}_n^X(W_i) &= \tfrac{n-1}{n(n+1)}(i + \tfrac{1}{n-1} + \tfrac{W_i - W_{i-1}}{W_{i+1} - W_{i-1}}) \\ &= \hat{F}_n^W(X_i),\end{aligned}$$

for $i = 1, ..., n$. So we have

$$H_a^W(m,n) = \frac{1}{n}\sum_{i=1}^{n} \ln \frac{k(X_{i+m} - X_{i-m})}{\hat{F}_n(kX_{i+m}) - \hat{F}_n(kX_{i-m})} = \ln k + H_a^X(m,n). \qquad \square$$

Theorem 3.4. *Under the assumptions of Theorem 3.3 the following relations hold.*

1. $E(H_b^W(m,n)) = \ln k + E(H_b^X(m,n))$
2. $Var(H_b^W(m,n)) = Var(H_b^X(m,n))$
3. $MSE(H_b^W(m,n)) = MSE(H_b^X(m,n))$

Proof. The proof is similar to the proof of Theorem 3.3 and is, thus, omitted. $\square$

4. Simulation Results

In this section we compare the bias and root mean squared error of our proposed estimator with those of the estimators of Ebrahimi et al.(1994)'s and those of the estimator based on kernel density estimate of f.
As can be seen from the tables, in almost all cases the bias and RMSE of the proposed estimators are less than those of Ebrahimi et al.(1994)'s and the estimator based on the kernel density estimator.
The output of kernel density of R 2.0.0 software, which was used for the kernel density estimate, has 512 ordered pairs $(x, \hat{f}_d)$, in which $\hat{f}_d$ is the value of the density estimate at point x.
The entropy estimator based on kernel density is thus

$$H_k = -\sum_{i=1}^{512} \ln(\hat{f}_{di})\hat{f}_{di}\Delta x,$$

where Δx is the difference between any two consecutive values of x in the output of kernel density estimate. The values in these tables are based on 10000 random samples, each of sizes 10, 20 and 30 and give the bias and root mean squared error of the estimators, for Uniform, Exponential and Normal distributions. It can be

Table 1. Bias and RMSE of entropy estimators for Uniform(0,1) distribution, $H(X) = 0$.

		Bias			RMSE		
n	m	H_c	H_d	H_b	H_c	H_d	H_b
10	1	-0.381	-0.250	-0.201	0.445	0.288	0.245
	2	-0.218	-0.151	-0.076	0.283	0.184	0.128
	3	-0.167	-0.122	-0.034	0.237	0.156	0.101
	4	-0.141	-0.108	-0.0126	0.215	0.143	0.093
	5	-0.127	-0.101	-0.001	0.209	0.140	0.095
20	1	-0.327	-0.261	-0.224	0.355	0.282	0.247
	2	-0.174	-0.141	-0.098	0.202	0.158	0.121
	3	-0.127	-0.105	-0.057	0.157	0.122	0.084
	4	-0.101	-0.085	-0.035	0.133	0.103	0.066
	5	-0.088	-0.075	-0.022	0.122	0.093	0.059
	6	-0.078	-0.067	-0.013	0.112	0.086	0.054
	7	-0.072	-0.063	-0.007	0.108	0.083	0.054
	8	-0.067	-0.059	-0.002	0.105	0.081	0.055
	9	-0.064	-0.057	0.001	0.106	0.081	0.057
	10	-0.061	-0.055	0.002	0.104	0.080	0.058
30	1	-0.307	-0.263	-0.230	0.325	0.276	0.245
	2	-0.158	-0.136	-0.106	0.175	0.147	0.120
	3	-0.112	-0.098	-0.065	0.130	0.109	0.081
	4	-0.089	-0.078	-0.043	0.108	0.090	0.062
	5	-0.075	-0.066	-0.031	0.096	0.078	0.051
	6	-0.067	-0.060	-0.023	0.088	0.072	0.046
	7	-0.060	-0.054	-0.016	0.083	0.067	0.043
	8	-0.056	-0.050	-0.012	0.079	0.064	0.042
	9	-0.051	-0.047	-0.008	0.076	0.061	0.042
	10	-0.049	-0.044	-0.005	0.074	0.060	0.041
	11	-0.047	-0.043	-0.003	0.073	0.059	0.041
	12	-0.045	-0.041	-0.002	0.073	0.058	0.041
	13	-0.043	-0.040	-0.000	0.073	0.059	0.043
	14	-0.041	-0.038	0.001	0.073	0.059	0.044
	15	-0.040	-0.037	0.002	0.073	0.058	0.045

seen from the above tables that bias and RMSE of H_b, H_c and H_d depend on the value of m. The optimum values of m for H_b for different sample sizes are

n	m
10	2
20	3
30	4

which do not depend on the true distribution, while the optimum value of m for H_c and H_d depend on the true distribution as well as on the sample size.

Table 2. Bias and RMSE of entropy estimators for Exponential distribution, $H(X) = 1$.

		Bias			RMSE		
n	m	H_c	H_d	H_b	H_c	H_d	H_b
10	1	-0.408	-0.203	-0.152	0.566	0.406	0.382
	2	-0.246	-0.0729	0.004	0.434	0.342	0.334
	3	-0.181	-0.002	0.088	0.401	0.336	0.347
	4	-0.130	0.062	0.160	0.383	0.345	0.375
	5	-0.095	0.113	0.215	0.385	0.370	0.413
20	1	-0.341	-0.241	-0.203	0.433	0.347	0.321
	2	-0.189	-0.106	-0.0627	0.310	0.257	0.242
	3	-0.134	-0.050	-0.001	0.279	0.241	0.235
	4	-0.099	-0.009	0.042	0.262	0.232	0.235
	5	-0.072	0.025	0.079	0.254	0.233	0.245
	6	-0.046	0.059	0.114	0.249	0.240	0.259
	7	-0.020	0.094	0.150	0.248	0.252	0.278
	8	0.008	0.132	0.189	0.254	0.273	0.305
	9	0.0309	0.164	0.222	0.266	0.296	0.331
	10	0.0564	0.197	0.256	0.274	0.318	0.357
30	1	-0.317	-0.254	-0.221	0.381	0.325	0.299
	2	-0.170	-0.118	-0.087	0.260	0.225	0.210
	3	-0.120	-0.068	-0.034	0.229	0.201	0.192
	4	-0.087	-0.032	0.003	0.212	0.191	0.188
	5	-0.065	-0.005	0.031	0.205	0.187	0.190
	6	-0.044	0.020	0.058	0.200	0.189	0.196
	7	-0.024	0.0449	0.083	0.197	0.192	0.204
	8	-0.004	0.071	0.109	0.201	0.202	0.219
	9	0.012	0.092	0.132	0.202	0.212	0.232
	10	0.030	0.116	0.156	0.205	0.222	0.245
	11	0.049	0.140	0.180	0.211	0.238	0.263
	12	0.068	0.164	0.204	0.221	0.254	0.282
	13	0.083	0.185	0.225	0.229	0.269	0.299
	14	0.098	0.205	0.246	0.240	0.287	0.317
	15	0.121	0.233	0.274	0.253	0.309	0.341

A serious drawback of the estimators proposed by Ebrahimi et al.(1994) is that the optimum values of m for their estimators very much depend on the type of the true distribution. The fact that in our procedure the optimum values of m are, more or less, the same for all types of distributions, is a clear advantage.

Acknowledgments

The second author acknowledges the support received from statistics center of Excellence at Ferdowsi University of Mashhad.

Table 3. Bias and RMSE of entropy estimators for Standard Normal distribution, $H(X) = 1.419$.

		Bias			RMSE		
n	m	H_c	H_d	H_b	H_c	H_d	H_b
10	1	-0.458	-0.038	-0.127	0.556	0.280	0.304
	2	-0.332	0.143	0.013	0.433	0.297	0.254
	3	-0.302	0.223	0.051	0.403	0.337	0.260
	4	-0.296	0.270	0.076	0.395	0.368	0.261
	5	-0.295	0.299	0.085	0.390	0.387	0.267
20	1	-0.363	-0.147	-0.189	0.419	0.242	0.269
	2	-0.227	0.029	-0.050	0.295	0.178	0.183
	3	-0.184	0.109	0.003	0.259	0.206	0.168
	4	-0.171	0.162	0.034	0.247	0.235	0.171
	5	-0.159	0.206	0.052	0.241	0.265	0.179
	6	-0.156	0.242	0.069	0.237	0.294	0.181
	7	-0.156	0.270	0.081	0.239	0.318	0.185
	8	-0.155	0.294	0.087	0.237	0.336	0.189
	9	-0.153	0.314	0.092	0.234	0.355	0.190
	10	-0.155	0.330	0.099	0.235	0.369	0.195
30	1	-0.329	-0.184	-0.211	0.369	0.242	0.261
	2	-0.194	-0.016	-0.079	0.244	0.144	0.161
	3	-0.150	0.056	-0.024	0.208	0.151	0.142
	4	-0.131	0.107	0.005	0.192	0.174	0.137
	5	-0.117	0.143	0.027	0.184	0.198	0.137
	6	-0.109	0.178	0.045	0.182	0.224	0.141
	7	-0.100	0.205	0.057	0.175	0.245	0.145
	8	-0.097	0.235	0.069	0.173	0.271	0.152
	9	-0.091	0.258	0.081	0.172	0.290	0.158
	10	-0.088	0.278	0.087	0.171	0.309	0.161
	11	-0.090	0.296	0.094	0.171	0.325	0.165
	12	-0.090	0.313	0.100	0.175	0.340	0.167
	13	-0.091	0.326	0.103	0.176	0.353	0.171
	14	-0.091	0.341	0.109	0.174	0.366	0.173
	15	-0.091	0.346	0.109	0.176	0.371	0.174

Table 4. Bias and RMSE of entropy estimators based on kernel density.

n	uniform(0,1)		Exponential		Normal(0,1)	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
10	0.175	0.273	0.112	0.412	-0.06	0.285
20	0.204	0.232	0.190	0.327	-0.001	0.177
30	0.203	0.218	0.204	0.296	0.019	0.139

References

1. I.A. Ahmad and P.E. Lin, A nonparametric estimation of the entropy for absolutely continuous distributions,*IEEE Trans, Inform. Theory* IT-22,327-375 (1976).

2. N. Ebrahimi, M. Habibullah, and E. Soofi, Testing exponentiality based on Kullback-Leibler Information, *JRSS Ser. B* 54,739-748 (1992).
3. N. Ebrahimi, K. Pflughoeft, and E. Soofi, Two measures of sample entropy,*Stat. Prob. Letters* 20, 225-234 (1994).
4. H. Joe, Estimation of entropy and other functionals of a multivariate density,*Ann. Inst. Statist. Math.* 41,683-697 (1989).
5. S.P. Mack, A comparative study of entropy estimators and entropy-based goodness-of-fit test, *Ph.D.Dissertation, University of California,* Riverside (1988).
6. A. Mokkadem, Estimation of entropy and information of absolutely continuous random variables, *IEEE TRans. Inform. Theory* IT-35, 193-196 (1989).
7. O. Vasicek, A test normality based on sample entropy,*JRSS Ser B* 38, 54-59 (1976).

Results on Monotonicity of the Residual Rényi Entropy of Order Statistics and Record Values

S. Zarezadeh and M. Asadi

Department of Statistics, University of Isfahan

This paper explores some monotonicity properties of the residual Rényi entropy of some ordered random variables. The residual Rényi entropy of the kth order statistic from a continuous distribution function is represented in terms of the residual Rényi entropy of the kth order statistic from uniform distribution. The monotone behavior of the residual Rényi entropy of order statistic under various conditions is discussed. The analogues results regarding the residual Rényi entropy of record values are also investigated.

Keywords: Order statistic, Record value; Rényi entropy, Residual lifetime, Incomplete beta function, Incomplete gamma function.

1. Introduction

In information theory to measure the amount of information in a probability distribution function a well known criterion is the Shannon entropy (which is introduced by Shannon [1] and is also known as differential entropy). Let X be a non-negative continuous random variable with density function f, distribution function F and the survival function $\bar{F} = 1 - F$. The Shannon entropy of F, which we denote by H, is defined as

$$H(X) = -\int_0^\infty f(x) \log f(x) dx. \tag{1}$$

The Shannon entropy H plays a central role in the field of information theory. It is known that $H(X)$ measures the uniformity of F. The larger amount of $H(X)$, more difficult to predict an outcome of F. Rènyi [2] introduced a one parameter extension of Shannon entropy that is more flexible than Shannon entropy and has a wide range of applications in many fields. The Rényi entropy of X, which we denote by $H_\alpha(X)$, is defined as follows:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \int_0^\infty f^\alpha(x) dx$$

where $\alpha > 0$, $\alpha \neq 1$. It is easy to show that when α tends to 1, $H_\alpha(X)$ tends to Shannon entropy in (1). Several properties of the Rényi entropy are explored

by many authors. Among others, we refer to Rényi [2], Morales et al. [3], Song [4] and Nadarjah and Zografos [5].

Let the random variable X denote a duration such as the lifetime of a system. Usually in reliability theory and survival analysis, when the system is still operating at time t, random variable of interest is the residual lifetime of the system. The residual lifetime of the system, which we denote by X_t, is $X_t = X - t|X > t$ with density function

$$f(x;t) = \frac{f(x+t)}{\bar{F}(t)}, \quad x > 0.$$

Ebrahimi [6] argued that to measure the entropy of the system lifetime X at time t, the measure (1) is no longer an appropriate measure and one should obtain the entropy of X_t. Then he introduced the following time dependent entropy which measures the uncertainty of the residual lifetime X_t

$$H(X;t) := -\int_0^\infty f(x;t)\log f(x;t)dx.$$

In last two decades, several authors have studied the properties of $H(X;t)$; see, for example, [7]-[9].

Gupta and Nanda [10] introduced the concept of Rényi entropy of the residual lifetime (RRE) as follows:

$$H_\alpha(X;t) = \frac{1}{1-\alpha}\log\int_t^\infty \frac{f^\alpha(x)}{\bar{F}^\alpha(t)}dx \tag{2}$$

where $\alpha > 0$, $\alpha \neq 1$. Various properties of this measure have been investigated by Asadi et al. [11], Nanda and Paul [12] and Mahmoudi and Asadi [13].

Different kinds of ordered random data are appeared in statistical studies. Order statistics are an important kind of ordered data which are used in many branches of probability and statistics including characterization of probability distributions, analysis of censored samples, reliability analysis, goodness-of-fit tests, quality control etc. Assuming that $X_1, ..., X_n$ is a random sample from F, the order statistics corresponding to the sample is defined by the arrangement of $X_1, ..., X_n$ from the smallest to the largest, denoted as $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$. For a comprehensive review on the theory and applications of order statistics one can refer to Daivid and Nagaraja [14].

The other important concept of ordered random variables which arises in many areas of applications is the concept of record values. Consider a sequence of i.i.d. random variables $\{X_i\ ; i \geq 1\}$. An observation X_j is called an upper record value if its value exceeds of all previous observations. Thus, X_j is an upper record if $X_j > X_i$ for every $i < j$. Examples, in which the record values arise naturally, include industrial stress testing, meteorological analysis, hydrology, sporting and athletic events, economy etc. For more details about records and their applications, one may refer to Arnold et al. [15].

The wide scope of applications of order statistics and record values gives a strong motivation to study the information properties of them. Recently attempts have been made regarding this. Ebrahimi et al. [16] explored some properties of the Shannon entropy of the order statistics. Baratpour et al. [17] investigated some results on the Rényi entropy of the order statistics and record values.

The aim of the present paper is to study some monotonicity properties of RRE of order statistics and record values. In Section 2, we represent the RRE of order statistics $X_{k:n}$ from distribution function F in terms of RRE of order statistics of uniform distribution. We show that, under some mild conditions the RRE of minima and the maxima of a sample which correspond to the lifetime of a series and a parallel system, respectively, are monotone functions of the number of the components in the system. Using a counter example we show that the RRE of other order statistics $X_{k:n}$ is not necessary monotone function of n. In the sequel of Section 2, we study the monotone behavior of RRE of order statistics $X_{k:n}$ in terms of k. It is shown that the RRE of $X_{k:n}$ is not a monotone function of k in entire support of F. Section 3, investigates the monotonicity properties of RRE of record values. In this section we also prove that, under some mild conditions, the RRE of record values is monotone function of number of records in the sequence.

Before giving the results of the paper we would like to mention here that, because of the page limit, we have not given the proofs of the theorems and lemmas in the manuscript. The proofs can be obtained from the authors upon request.

Throughout the paper increasing (decreasing) means non-decreasing (non-increasing).

2. The residual Rényi entropy of order statistics

In this section we concentrate on the RRE of order statistics. First note that the density function $f_{k:n}(x)$ and the survival function $\bar{F}_{k:n}(x)$, $k = 1, ..., n$ of $X_{k:n}$ are, respectively, given by

$$f_{k:n}(x) = \frac{1}{B(k, n-k+1)}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x) \tag{3}$$

and

$$\bar{F}_{k:n}(x) = \sum_{i=0}^{k-1}\binom{n}{i}F^i(x)\bar{F}^{n-i}(x) \tag{4}$$

where

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx, \quad a > 0,\ b > 0.$$

It is well known that $\bar{F}_{k:n}(t)$ can also be represented as

$$\bar{F}_{k:n}(x) = \frac{\bar{B}_{F(x)}(k, n-k+1)}{B(k, n-k+1)} \tag{5}$$

where

$$\bar{B}_x(a,b) = \int_x^1 u^{a-1}(1-u)^{b-1}du, \quad 0 < x < 1.$$

$B(a,b)$ and $\bar{B}_x(a,b)$ are known in the literature as the beta and the incomplete beta functions, respectively (see, for example, [14]).

Notation: Throughout this section we use the notation $Y \sim \bar{B}_t(a,b)$ to show that Y has a distribution with density function

$$f_Y(y) = \frac{1}{\bar{B}_t(a,b)} y^{a-1}(1-y)^{b-1} \quad ; t \leq y \leq 1. \tag{6}$$

Remark 2.1. A $(n-k+1)$-out-of-n system, which is an important structure in reliability engineering, functions if and only if at least k components out of n components function. Let $X_1, X_2, ..., X_n$ denote the independent lifetimes of n components of such system. Then it is well known that the lifetime of a $(n-k+1)$-out-of-n system is equal to the kth order statistics $X_{k:n}$. The special cases of $k = 1$ and $k = n$ are corresponding to systems which are known as series and parallel systems, respectively.

The result of the following lemma is useful in the sequel.

Lemma 2.1. *Let $U_{k:n}$ be kth order statistic based on a sample of size n from uniform distribution on $(0,1)$. Then*

$$H_\alpha(U_{k:n};t) = \frac{1}{1-\alpha}\log \bar{B}_t(\alpha(k-1)+1, \alpha(n-k)+1) - \frac{\alpha}{1-\alpha}\log \bar{B}_t(k, n-k+1)$$

It is well known, from the probability integral transformation, that $U_{k:n} \overset{d}{=} F(X_{k:n}),\ k = 1, ..., n$ where $X_{k:n}$ is the kth order statistic based on a random sample of size n from continuous distribution F (see, for example, [14]). Using this, the RRE of order statistics $X_{k:n}$ from an absolutely continuous distribution can be represented in terms of RRE of order statistics of uniform distribution function. This is shown in the following theorem.

Theorem 2.1. *The RRE of the kth order statistic from an absolutely continuous distribution function F having density function f can be represented in terms of the RRE of kth order statistic from uniform distribution, over the unit interval, as follows*

$$H_\alpha(X_{k:n};t) = H_\alpha(U_{k:n};F(t)) + \frac{1}{1-\alpha}\log E[f^{\alpha-1}(F^{-1}(Y_k))]$$

where $Y_k \sim \bar{B}_{F(t)}(\alpha(k-1)+1, \alpha(n-k)+1)$.

Remark 2.2. The quantity $f(F^{-1}(x))$ is known, in the literature, as the density-quantile function and is used to approximate the moments of order statistics (see [14]).

Under the assumptions of Theorem 2.1, we have the following corollary regarding the residual Shannon entropy of order statistics.

Corollary 2.1. *The residual Shannon entropy of kth order statistic from absolutely continuous distribution function F can be written in terms of the residual Shannon entropy of kth order statistic from uniform distribution, over the unit interval, as follows*

$$H(X_{k:n};t) = H(U_{k:n};F(t)) - E[\log f(F^{-1}(Y_k))] \tag{7}$$

where $Y_k \sim \bar{B}_{F(t)}(k, n-k+1)$. The specialized version of this result for $t=0$, is already obtained by Ebrahimi et al. [16].

In the following we give one example.

Example 2.1. Suppose that X is a random variable having the exponential distribution with mean $\frac{1}{\theta}$. Then $f(F^{-1}(y)) = \theta(1-y)$ and we have

$$E[f^{\alpha-1}(F^{-1}(Y_1))] = \frac{\theta^{\alpha-1}\bar{F}^{n\alpha}(t)}{n\alpha\bar{B}_{F(t)}(1,\alpha(n-1)+1)}$$

For $k=1$, Theorem 2.1 gives

$$H_\alpha(X_{1:n};t) = \frac{\log\alpha}{\alpha-1} - \log(n\theta)$$

On the other hand, we have

$$H_\alpha(X;t) = \frac{\log\alpha}{\alpha-1} - \log\theta.$$

This gives

$$H_\alpha(X_{1:n};t) - H_\alpha(X;t) = -\log n.$$

That is, the difference between of RRE of the lifetime of a series system and RRE of the lifetime of each components is free of time and, in fact, is equal to $-\log n$, where n is the number of the components of the system.

In the following we explore some monotone behavior of RRE of order statistics. First we prove the following lemma.

Lemma 2.2. *Consider a series (parallel) system with n components each having uniform distribution over unit interval. Then the RRE of the system lifetime is a decreasing function of the number of components.*

We use the result of the lemma to prove the following theorem.

Theorem 2.2. *Let $X_1, ..., X_n$ be a set of i.i.d random variables representing the lifetime of the components of a series (parallel) system having a common distribution function F. Assume that F has a density function f which is decreasing (increasing) in its support. Then the RRE of system lifetime is decreasing in n.*

The following example shows that the result of above theorem is not in general valid for any $(n-k+1)$-out-of-n system.

Example 2.2. Assume that the structure of the system is $(n-1)$-out-of-n. Then the lifetime of the system is $X_{2:n}$. Assume that the components of the system have uniform distribution on $(0,1)$. Figure 1 shows the graph of RRE of $X_{2:n}$ at time $t = 0.2$ for $n = 2, \ldots, 30$. This is evident from the graph that the RRE of the system is not a decreasing function of n. In fact the graph shows that RRE of $X_{2:2}$ is less than that of $X_{2:3}$.

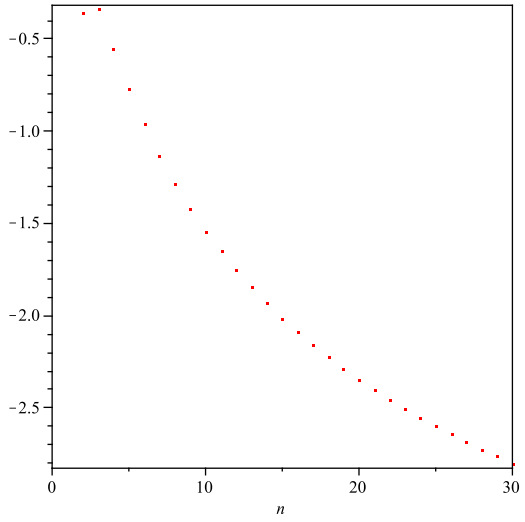

Fig. 1. The plot of RRE of $(n-1)$-out-of-n system, $n = 2, \ldots, 30$ when the parent distribution is uniform.

Remark 2.3. Another situation in which the RRE of first order statistic $X_{1:n}$, is decreasing in the number of the components of the system arises as follows. Asadi et al. [11] showed that when X and Y are two non-negative continuous random variables, under the condition that either X or Y has a decreasing failure rate, the $X \leq_{hr} Y$ (that is the hazard rate of X is more than the hazard rate of Y everywhere) implies that $H_\alpha(X;t) \leq H_\alpha(Y;t)$ for all $t \geq 0$. Now assuming that X has a decreasing failure rate, based on Asadi et al.'s result and the fact that $X_{1:n} \leq_{hr} X$, we have $H_\alpha(X_{1:n};t) \leq H_\alpha(X;t)$. If X has decreasing failure rate then $X_{1:n}$, $n \geq 1$, has also a decreasing failure rate. On the other hand

$$X_{1:1} \geq_{hr} X_{1:2} \geq_{hr} \ldots \geq_{hr} X_{1:n}$$

Hence, we have

$$H_\alpha(X_{1:1};t) \geq H_\alpha(X_{1:2};t) \geq ... \geq H_\alpha(X_{1:n};t).$$

That $H_\alpha(X_{1:n};t)$ is decreasing in n. Examples of distributions with decreasing failure rate include Weibull distributions with shape parameter less that one, Gamma distribution with shape parameter less than one, mixture of two exponential distributions, etc. Hence for these distributions the RRE of a series system is a decreasing function of the number of components.

In the following we study the behavior of RRE of order statistics $X_{k:n}$ in terms of k. First we prove the following lemma.

Lemma 2.3. *Let $U_{k:n}$ denote the kth order statistic of uniform distribution over unit interval. If $k_1 \leq k_2 \leq n$, are integers then for $t \geq \frac{k_2-1}{n-1}$,*

$$H_\alpha(U_{k_1:n};t) \leq H_\alpha(U_{k_2:n};t).$$

As a conclusion of the above lemma we have the following theorem.

Theorem 2.3. *Let X be a non-negative absolutely continuous random variable with distribution function F. Let F have a density function f which is decreasing in its support. If k_1 and k_2 are integers such that $k_1 \leq k_2 \leq n$, then $H_\alpha(X_{k_1:n};t) \leq H_\alpha(X_{k_2:n};t)$ for $t \geq F^{-1}(\frac{k_2-1}{n-1})$.*

The class of distribution functions with decreasing density functions are a wide class of distributions. Examples are exponential, Pareto, mixture of exponential and Pareto distributions etc.

The following example shows that the condition $t \geq F^{-1}(\frac{k_2-1}{n-1})$ can not be dropped from the conditions of theorem.

Example 2.3. Let the density function of X be given by

$$f(x) = \frac{(1-x)e^{(-\frac{4x}{3+3x})}}{(1+x)^{1/3}}, \quad 0 < x < 1.$$

Figure 2 displays the plots of RRE of order statistics $X_{k:7}$, for $k = 1, 2, ..., 7$ based on density function of X. It is seen from the plots that the RRE of the order statistics are not ordered in terms of k for all values of t, $t \in (0, 1)$.

Corollary 2.2. *Let X be a non-negative absolutely continuous random variable with decreasing density function f. If $k \leq \frac{n+1}{2}$ then $H_\alpha(X_{k:n};t)$ is increasing in k for values greater than the median of distribution.*

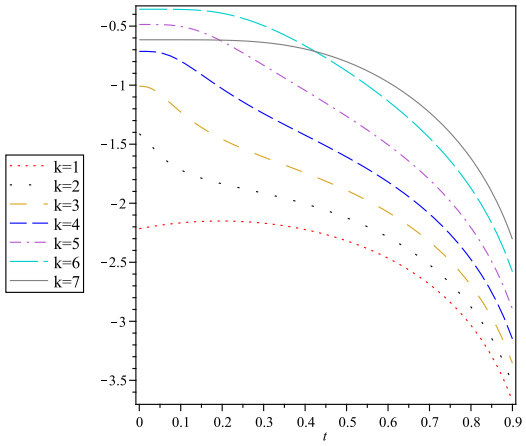

Fig. 2. The plot of RRE of $X_{k:7}$, $k = 1, ..., 7$ based on the parent distribution given in Example 2.3

3. The Residual Rényi entropy of record values

In this section we focus on the RRE of record values. Let $U_1, U_2, ...$ be a sequence of upper record values based on a sequence of non-negative continuous random variables X_i's with cdf F and pdf f. Then the density function and survival function of U_n, which we denote by F_{U_n} and f_{U_n}, respectively, are given by

$$f_{U_n}(x) = \frac{[-\log\ \bar{F}(x)]^{n-1}}{(n-1)!}\ f(x), \quad x > 0, \quad n \geq 1$$

$$\begin{aligned}\bar{F}_{U_n}(x) &= \sum_{j=0}^{n-1} \frac{[-\log\ \bar{F}(x)]^j}{j!}\ \bar{F}(x) \\ &= \frac{\Gamma(n; -\log \bar{F}(x))}{\Gamma(n)}.\end{aligned}$$

The result of the following lemma is easy to verify.

Lemma 3.1. *Let U_n^* denote the nth upper record value from $U(0,1)$. Then*

$$H_\alpha(U_n^*; t) = \frac{1}{1-\alpha} \log \frac{\Gamma(\alpha(n-1)+1; -\log(1-t))}{\Gamma^\alpha(n; -\log(1-t))}.$$

The following theorem represents the RRE of upper record U_n in terms of upper record U_n^* of uniform distribution.

Theorem 3.1. *The RRE of U_n can be written in terms of the RRE of U_n^* as follows*

$$H_\alpha(U_n;t) = H_\alpha(U_n^*;F(t)) + \frac{1}{1-\alpha}\log E[f^{\alpha-1}(F^{-1}(1-e^{-V_n}))]$$

where $V_n \sim \Gamma_{-\log \bar{F}(t)}(\alpha(n-1)+1,1)$.

Under assumptions of Theorem 3.1, we have the following corollary regarding the residual Shannon entropy of upper record values.

Corollary 3.1. *The residual Shannon entropy of nth upper record value of continuous distribution function F can be written in terms of the residual Shannon entropy of nth upper record value of $U(0,1)$ as follows*

$$H(U_n;t) = H(U_n^*;F(t)) - E[\log f(F^{-1}(1-e^{-Z_n}))] \tag{8}$$

where $Z_n \sim \Gamma_{-\log \bar{F}(t)}(n,1)$.

Example 3.1. Let X have Weibull distribution with density

$$f(x) = \beta\lambda^\beta(x-\mu)^{\beta-1}e^{-[\lambda(x-\mu)]^\beta} \quad , x \geq \mu.$$

Here, $F^{-1}(x) = \frac{1}{\lambda}(-log(1-x))^\beta + \mu$. Then we have for $\beta \geq 1$,

$$E[f^{\alpha-1}(F^{-1}(1-e^{-V_n}))] = \frac{(\lambda\beta)^{\alpha-1}}{\Gamma(\alpha(n-1)+1;(\lambda(t-\mu))^\beta)}\frac{\Gamma(\frac{1}{\beta}(1-\alpha)+n\alpha;\alpha(\lambda(t-\mu))^\beta)}{\alpha^{\frac{1}{\beta}(1-\alpha)+n\alpha}}.$$

Therefore

$$H_\alpha(U_n;t) = \frac{1}{1-\alpha}\log\frac{\Gamma(\frac{1}{\beta}(1-\alpha)+n\alpha;\alpha(\lambda(t-\mu))^\beta)}{\Gamma^\alpha(n;(\lambda(t-\mu))^\beta)} - \log(\lambda\beta) - \frac{1}{\beta}\log\alpha - \frac{n\alpha}{1-\alpha}\log\alpha.$$

The following lemma and theorem investigate the behavior of RRE of upper records in terms of n.

Lemma 3.2. *The RRE of upper record values of uniform distribution on (0,1) is decreasing in n.*

Now we can prove the following theorem.

Theorem 3.2. *Let $\{X_i, i \geq 1\}$ be a sequence of i.i.d random variables from distribution function F having an increasing density function f. If $\{U_n, n \geq 1\}$ represents the sequence of upper record values corresponding to F, then $H_\alpha(U_n;t)$ is decreasing in n.*

An example of the distributions for which this theorem can be applied is the power distribution with distribution function $F(x) = x^\beta$, $0 < x < 1$, $\beta > 1$.

References

1. C.E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* 27 (1948) 379-423.
2. A. Renyi, On Measures of entropy and information, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, University of California Press, 1961, pp. 547-561.
3. D. Morales, L. Pardo, I. Vajda, Some new statistics for testing hypotheses in parametric models, *Journal of Multivariate Analysis* 62 (1997) 137-168.
4. K. Song, Rényi information, loglikelihood and an intrinsic distribution measure, *Journal of Statistical Planning and Inference* 93 (2001) 51-69.
5. S. Nadarjah, K. Zografos, Formulas for Rényi information and related measures for univariate distributions, *Inform. Sci.* 155 (2003) 119-138.
6. N. Ebrahimi, How to measure uncertainty in the residual lifetime distributions, *Sankhya A* 58 (1996) 48-57.
7. N. Ebrahimi, S.N.U.A. Kirmani, Some results on ordering of survival functions through uncertainty, *Statistics and Probability Letters* 11 (1996) 167-176.
8. M. Asadi, N. Ebrahimi, Residual entropy and its characterizations in terms of hazard function and mean residual life function, *Statistics and Probability Letters* 49 (2000) 263-269.
9. F. Belzunce, J. Navarro, J.M. Ruiz, Some results on residual entropy functions, *Metrika* 59 (2004) 147-161.
10. R.D. Gupta, A.K. Nanda, α- and β-Entropies and relative entropies of distributions, *Journal of Statistical Theory and Applications* 1(3) (2002) 177190.
11. M. Asadi, N. Ebrahimi, E.S. Soofi, Dynamic generalized information measures, *Statistics and Probability Letters* 71 (2005) 85-98.
12. A.K. Nanda, P. Paul, Some results on generalized residual entropy, *Information Sciences* 176 (2006) 2747.
13. M. Mahmoudi, M. Asadi, More on the residual and past entropies of order α, Submitted for publication, (2010).
14. H.A. David, H.N. Nagaraja, *Order Statistics*, third ed. (Wiley, New York, 2003).
15. B.C. Arnold, N. Balakrishnan, H.N. Nagaraja, *Records*. (Wiley, New York, 1998).
16. N. Ebrahimi, E.S. Soofi, H. Zahedi, Information properties of order statistics and spacings, *IEEE Transactions on Information Theory* IT50 (2004) 177-183.
17. S. Baratpur, J. Ahmadi, N.R. Arghami, Characterizations based on Rényi entropy of order statistics and record values, *Journal of Statistical Planning and Inference* 138 (2008) 2544-2551.

AUTHOR INDEX