Proceedings
of the
Seventh Iranian Statistical Conference

Contributed Papers
(Volume 3)

August 23-25, 2004

First Published

Allame Tabataba'i University
Tehran 2006
ISBN 964 – 8415 – 40 – 4

Publication no.: 193

ISBN 964 – 8415 – 40 – 4

Proceedings of the Seventh Iranian Statistical Conference

Contributed Papers

(Volume 3)

August 23-25, 2004

First Published: 2006 Circulation: 800
Allame Tabataba'i University
Tehran, Iran, P.O. Box 15815/3487

Preface

The Seventh Iranian Statistical Conference held at Allameh Tabatabaie University on August 23-25 2004.

This Proceedings volume includes contributed papers presented at the Seventh Iranian Statistical Conference(ISC7). The papers in this volume are organized according to the last name of the first author(s) of the papers.

The scientific and organizing committees would like to thank Y. Akbari and S. Mohandesi for typesetting, students of Statistics Department at Allameh Tabatabaie University for valuable helps and Allameh-Tabatabaie University for printing and production of this proceedings volume.

Scientific and organizing Committees of ISC7
September 2006

Organizing Committee (Allameh Tabatabaie University)

Asghari Oskoie, M.R.
Badamchizadeh A.
Bameni Moghadam M.
Eskandari F.
Mirian, S.A.
Navabpour, H.R.
Neisy, A.
Nematollahi, N. (Chairman)
Salili, Sh.
Zali, M.J.

Scientific Committee

Eskandari, F.	Allameh Tabatabaei University
Gerami, A.	Statistical Research Center
Jelodari Mamaghani, M.	Allameh Tabatabaie University
Meshkani, M.R.	Shahid Beheshti University
Mehrabi, Y.	Shahid Beheshti University
Mohammadzadeh, M.	Tarbiat Modarres University
Navabpour, H.R. (Chairman)	Allameh Tabatabaie University
Nematollahi,N.	Allameh Tabatabaie University
Parsian, A.	Isfahan University of Technology
Shams, S.	Alzahra University
Vahidi Asl, M. Q.	Shahid Beheshti University

Contents

Bounds and approximations for the moments of record statistics by using wavelets

J. Ahmadi and M. Fashandi

School of Mathematical Sciences, Ferdowsi University, Mashhad, 91775 Iran

Abstrcat: Let $\{X_i,\ i \geq 1\}$ be a sequence of iid continuous random variables. 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 all $i < j$. In this paper, first, based on a Taylor expansion, an approximation for the moments of record values in terms of the quantile function and its derivatives is derived. Then, we present the wavelet expansion method for deriving bounds as well as approximations for the means of record values. Some tables are provided for comparing the results obtained by the method of this paper with the exact values.

Keywords: Record values, Wavelet expansion, Taylor expansion, Quantiles, Cauchy-Schwarz inequality, Haar Wavelet.

1 Introduction

Let $\{X_i,\ i \geq 1\}$ be a sequence of iid continuous random variables, each distributed according to a cumulative distribution function (cdf) F, and probability density function (pdf) f. An observation X_j is 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 all $i < j$. An analogous definition can be given for lower record values.

Record values can be viewed as order statistics from a sample whose size is determined by the values and the order of occurrence in the observations. These statistics are of interest and importance in several real-life data. Like extreme order statistics, record values are applied in estimating strength of materials, predicting natural disasters and sport results etc, for more details on applications of the record values, see Arnold et al. (1998), Ahmadi (2000).

Several authors have studied of bounds for the moments and approximations of order statistics. In the context of record values, Nagaraja (1978) presented analytic formulas for sharp bounds of record statistics, by using Schwarz inequality (see, also, Arnold and Balakrishnan (1989)). By the same approach, Ahmadi and Arghami (2001) obtained some bounds for the differences of record values. Raqab (2000) using the Holder's inequality and derived some bounds for expectations of record statistics. Gajek and Okolewski (2000) applied the Steffensen inequality to obtain different bounds for expectations of order and record statistics. Raqab and Rychik (2002) presented sharp bounds for the mean of the k-th record value using the Moriguti monotone approximations combined with the Holder's inequality.

Wavelet Analysis is a particular time-scale representation of signals that has become popular in mathematics, physics and engineering in the last few years. This theory has been investigated extensively in both theory and applications. Many interesting papers on this topic have been published. The beautiful theory of wavelets have many applications in sciences including signal analysis (see, Cheney and Light (2000)), reconstruction of images (see, Mallat, (1989)), and seismic processing (see, Alkemade, (1995)).

In this paper, Section 2 contains some preliminaries. In Section 3, we develop some series approximations for the moments of record values and present in particular, the formulas for the means of record statistics. In Section 4, we develop the wavelet expansion method for deriving bounds as well as approximations for the mean of record values. Some tables are provided for comparison the results of this paper with the exact values. Also, In Section 5, we present the Maple procedure for obtaining the numerical results of this paper.

2 Preliminaries

For the convenience of notation, we shall denote the i-th upper record value by U_i. The zero-th upper being $U_0 = X_1$, which is referred to as the reference value or the trivial record. Let f_r denote the pdf of U_r. The marginal density of U_r is given by

$$f_r(u_r) = \frac{[-\log(1-F(u_r))]^r}{r!} f(u_r). \qquad (2-1)$$

and the expectation of U_r is given by

$$E(U_r) = \int_{-\infty}^{+\infty} x \frac{[-\log(1-F(x))]^r}{r} f(x)dx.$$

By using transformation $u = F(x)$, we can write

$$E(U_r) = \int_0^1 F^{-1}(u) \frac{[-\log(1-u))]^r}{r} du \qquad (2-2)$$

For the special case when X_i's come from standard uniform $(0,\ 1)$, we have the following lemma.

Lemma 1: Let V_r be the n-th upper record value from $U(0,\ 1)$, then
i) $E(V_r) = 1 - \frac{1}{2^{r+1}}$,
ii) $E\left(V_r - p_r\right)^j = \sum_{i=0}^{j} \binom{j}{i} \frac{(-1)^i q_r^{j-i}}{(1+i)^{r+1}}$,
where $p_r = E(V_r)$ and $q_r = 1 - p_r$.
Proof: By (2-1) and (2-2) we have

$$\begin{aligned} E(V_r) &= \int_0^1 v \frac{\left[-\log(1-v)\right]^r}{r!} dv \\ &= \int_0^{+\infty} (1 - e^{-y}) \frac{y^r}{r!} e^{-y} dy \\ &= 1 - \frac{1}{2^{r+1}}. \end{aligned} \tag{2-3}$$

Also, by (2-2) and (2-3), we get

$$\begin{aligned} E\left(V_r - p_r\right)^j &= \int_0^1 (v - p_r)^j \frac{\left[-\log(1-v)\right]^r}{r!} dv \\ &= \int_0^{+\infty} (q_r - e^{-y})^j \frac{y^r}{r!} e^{-y} dy \\ &= \sum_{i=0}^{j} \binom{j}{i} (-1)^i q_r^{j-i} \int_0^{+\infty} \frac{y^r}{r!} e^{-(i+1)y} dy \\ &= \sum_{i=0}^{j} \binom{j}{i} \frac{(-1)^i q_r^{j-i}}{(1+i)^{r+1}}. \end{aligned} \tag{2-4}$$

3 Taylor series

In this section, based on a Taylor series expansion, we obtain reasonable approximations for the moments of record values. Similar approximations for the moments of order statistics are presented in David and Nagaraja (2003) Chapter 4, page 60 and Arnold et al. (1992) Chapter 5, page 128. Childs and Balakrishnan (2002) provide Maple procedures facilitating the computations and permitting the inclusion of additional terms.

As pointed out in Lemma 1, the transformation $Y = F(X)$ takes the continuous record values U_r to the n-th record value V_r of the standard uniform $(0, 1)$ distribution. We now invert the relation $V_r = F(U_r)$ by writing

$$U_r = Q(V_r), \tag{3-1}$$

where $Q(.) = F^{-1}(.)$. By expanding $Q(V_r)$ in a Taylor series around the point $E(V_r) = 1 - 1/2^{n+1} = p_r$, we get the following series expansion for U_r

$$\begin{aligned} U_r &= Q(V_r) = Q(p_r) + (V_r - p_r)Q^{(1)}(p_r) + \frac{1}{2!}(V_r - p_r)^2 Q^{(2)}(p_r) \\ &+ \frac{1}{3!}(V_r - p_r)^3 Q^{(3)}(p_r) + \frac{1}{4!}(V_r - p_r)^4 Q^{(4)}(p_r) + \cdots , \qquad (3-2) \end{aligned}$$

where $Q^{(1)}(p_r), Q^{(2)}(p_r), Q^{(3)}(p_r), Q^{(4)}(p_r), ...$ are the successive derivatives of $Q(v) = F^{-1}(v)$ evaluated at $v = p_r$.

Now, by taking expectation on both sides of (3-2) and using the expressions of (2-3), we derive

$$E(U_r) \propto \sum_{j=0}^{k} \frac{Q^{(j)}(p_r)}{j!} \sum_{i=0}^{j} \binom{j}{i} \frac{(-1)^i q_r^{j-i}}{(1+i)^{r+1}}, \qquad (3-3)$$

where $q_r = 1 - p_r = 1/2^{n+1}$. If, we consider the first four terms of (3-3) and replacing $Q(p_r)$ by Q_r , after some simple calculations we obtain

$$\begin{aligned} E(U_r) &\simeq q_r^4 \{ \frac{Q_r^{(4)}}{4!} \} + q_r^3 \{ \frac{Q_r^{(4)}}{3! \times 2^{r+1}} - \frac{Q_r^{(3)}}{3!} \} \\ &+ q_r^2 \{ \frac{Q_r^{(4)}}{4! \times 3^{r+1}} - \frac{Q_r^{(3)}}{2 \times 3^{n+1}} + \frac{Q_r^{(2)}}{2!} \} \\ &+ q_r \{ \frac{Q_r^{(4)}}{3! \times 4^{r+1}} - \frac{Q_r^{(3)}}{2! \times 3^{r+1}} + \frac{Q_r^{(4)}}{2!} - Q_r^{(1)} \} \\ &+ \{ \frac{Q_r^{(4)}}{4! \times 5^{r+1}} - \frac{Q_r^{(3)}}{3! \times 4^{r+1}} + \frac{Q_r^{(2)}}{2! \times 3^{r+1}} - \frac{Q_r^{(1)}}{2!} + Q_r) \}. \qquad (3-4) \end{aligned}$$

The evaluation of the derivatives of $Q(v) = F^{-1}(v)$ is rather straightforward since $F^{-1}(v)$ being available explicitly (see Arnold et al. (1992). For example, for the logistic distribution with cdf

$$F(x) = \frac{1}{1+e^{-x}}, -\infty < x < +\infty, \qquad (3-5)$$

we have $x = Q(v) = F^{-1}(v) = \log v - \log(1-v)$. Then

$$\begin{aligned} Q^{(1)}(v) &= \frac{1}{v} + \frac{1}{1-v}, \\ Q^{(2)}(v) &= -\frac{1}{v^2} + \frac{1}{(1-v)^2}, \\ Q^{(3)}(v) &= 2\{ \frac{1}{v^3} + \frac{1}{(1-v)^3} \}, \\ Q^{(4)}(v) &= 6\{ -\frac{1}{v^4} + \frac{1}{(1-v)^4} \}, \qquad (3-6) \end{aligned}$$

and so on. For the purpose of illustration, we consider here the logistic distribution given by (3-5). By using the values in (3-6), we computed approximations for $E(U_r)$ from (3-3) for $r = 1$ up to 7 with $k = 1$ up to 4. These values are presented in Table I along with the exact values. We used the Mathematical Package *Maple 8,* for numerical results in this paper.

Table I. Approximations in (3-3) for $E(U_r)$ from the logistic distribution

r \ k	0	1	2	3	4	Exact Value
1	1.0986	1.0986	1.4938	1.5705	1.7307	1.6449
2	1.9459	1.9459	2.5264	2.4769	2.7810	2.8470
3	2.7080	2.7081	3.4945	3.1324	3.9531	3.9293
4	3.4339	3.4340	4.4864	3.4109	6.0996	4.9662
5	4.1431	4.1431	5.5474	2.9283	12.14879	5.9836
6	4.8441	4.8442	6.7169	0.8544	32.2489	6.9919
7	5.5412	5.5413	8.0384	-4.5190	100.5953	7.9960

Fortunately, the evaluation of the derivatives of $Q(v) = F^{-1}(v)$ is not difficult even in the case of distribution where $Q(v)$ does not exist in an explicit form. In this case, by using the fact

$$Q'(v) = \frac{1}{f(Q(v))}, \qquad (3-7)$$

where $f(Q(v))$ is the pdf of X evaluated at $Q(v)$. One can derive the higher-order derivations of $Q(v)$ by successively differentiating the expression of $Q'(v)$ in (3-7). For example in the standard normal family with cdf $\Phi(v)$ and pdf $\phi(v)$ (see, David and Nagarja (2003)), we have $Q(v) = \Phi^{-1}(v)$, then

$$\begin{aligned} Q^{(2)}(v) &= \frac{Q(v)}{\phi^2(Q(v)}, \\ Q^{(3)}(v) &= \frac{1+2Q^2(v)}{\phi^3(Q(v))}, \\ Q^{(4)}(v) &= \frac{Q(7+6Q^2(v))}{\phi^4(Q(v)}, \end{aligned}$$

since $(d/dx)\phi(x) = -x\phi(x)$.

4 Wavelets

The Haar Wavelet is defined as follows:
Definition 1: The Haar wavelet ψ is defined by

$$\psi(x) = 1_{[0,1/2)}(x) - 1_{[1/2,1)}(x), \qquad (4-1)$$

in which 1_A is the indicator function of the set A.

It is easy to see that, ψ is 0 outside of $[0,1)$, and

$$\int_{-\infty}^{+\infty} \psi(x)dx = 0, \qquad \int_{-\infty}^{+\infty} \psi^2(x)dx = 1.$$

It is a well-known fact that the set $\{2^{m/2}\psi(2^m x - n), m, n \in Z\}$ is an orthonormal basis for $L^2(R)$. For simplicity, we use the following notation throughout this paper:

$$\psi_{m,n}(x) = 2^{m/2}\psi(2^m x - n).$$

In this work, we need to expand a function in $L^2[0,1)$, in its wavelet series form. So we should find a basis for $L^2[0,1)$. Using Haar wavelet and periodizing the Haar system with a set of periodic wavelet, with period 1, we obtain:

$$\widetilde{\psi}_{m,n}(x) = \sum_{j\in Z} 2^{m/2}\psi(2^m(x+j) - n).$$

Theorem 1: The function 1 and $\widetilde{\psi}_{m,n}$, for $m \geq 0$ and $n = 0, 1, ..., 2^n - 1$ form an orthonormal basis for $L^2[0,1)$.

By Theorem 1, each $f \in L^2[0,1)$ could be written in the from: (see, Walker (1997))

$$f(x) = A_{0,0} + \sum_{m=0}^{+\infty} \sum_{n=0}^{2^m-1} \alpha_{m,n}\widetilde{\psi}_{m,n}(x),$$

where

$$A_{0,0} = \int_0^1 f(x)dx, \qquad (4-2)$$

$$\alpha_{m,n} = \int_0^1 f(x)\widetilde{\psi}_{m,n}(x)dx; \quad m \geq 0, n = 0, 1, ..., 2^m - 1. \qquad (4-3)$$

Thus, we can write

$$f(u) = A_{00} + \lim_{k\to\infty} \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \alpha_{m,n}\widetilde{\psi}_{m,n}(u).$$

Similarly, for $g \in L^2[0,1)$ we have

$$g(u) = B_{0,0} + \sum_{m=0}^{+\infty} \sum_{n=0}^{2^m-1} \beta_{m,n}\widetilde{\psi}_{m,n}(u),$$

where

$$B_{0,0} \quad = \quad \int_0^1 g(x)dx, \qquad (4-4)$$

$$\beta_{m,n} \quad = \quad \int_0^1 g(x)\widetilde{\psi}_{m,n}(x)dx; \quad m \geq 0, n = 0, 1, ..., 2^m - 1. \qquad (4-5)$$

Then

$$g(u) = B_{00} + \lim_{k \to \infty} \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \beta_{m,n}\widetilde{\psi}_{m,n}(x).$$

For any two functions $f(u), g(u) \in L^2[0,1)$ with wavelet coefficients $\{\alpha_{m,n}, m \geq 0, n = 0, 1, ..., 2^n - 1\}$ and $\{\beta_{m,n}, m \geq 0, n = 0, 1, ..., 2^n - 1\}$, respectively, we may observe that for any $k \geq 0$

$$\begin{aligned}
& |\int_0^1 f(u)g(u)du - (A_{0,0}B_{0,0} + \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \alpha_{m,n}\beta_{m,n})| \\
= \quad & |\int_0^1 \left\{ f(u) - (A_{00} + \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \alpha_{m,n}\widetilde{\psi}_{m,n}(u)) \right\} \\
\times \quad & \left\{ g(u) - (B_{00} + \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \beta_{m,n}\widetilde{\psi}_{m,n}(u)) \right\} du \\
\leq \quad & \left[\left\{ \int_0^1 f^2(u)du - (A_{0,0}^2 + \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} lpha_{m,n}^2) \right\} \right]^{1/2} \\
\times \quad & \left[\left\{ \int_0^1 g^2(u)du - (B_{0,0}^2 + \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \beta_{m,n}^2) \right\} \right]^{1/2}
\end{aligned}$$

where the last inequality obtains by applying the Cauchy-Schwarz inequality, in which

$$f(u) = F^{-1}(u) \quad \text{and} \quad g(u) = \frac{1}{r!}\{-\log(1-u)\}^r.$$

From (4-2), (4-4) and for any arbitrary continuous distribution function with mean 0 and variance 1, we get

$$A_{0,0} = \int_0^1 f(u)du = 0, \quad \int_0^1 f^2(u)du = 1,$$

$$B_{0,0} = \int_0^1 g(u)du = 1 \;\text{ and }\; \int_0^1 g^2(u)du = \binom{2r}{r}.$$

Then, we immediately obtain

$$|\mu_r - \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \alpha_{m,n}\beta_{m,n}| \\ \leq \left[\left\{1 - \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \alpha^2_{m,n}\right\}\left\{\binom{2r}{r} - (1 + \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \beta^2_{m,n})\right\}\right]^{1/2} . \quad (4\text{-}6)$$

In the above inequality, the equality holds if and only if

$$F^{-1}(u) - \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \alpha_{m,n}\widetilde{\psi}_{m,n}(u) \propto \frac{1}{r!} - \log(1-u)^r - \sum_{m=0}^{k} \sum_{n=0}^{2^m-1} \alpha_{m,n}\widetilde{\psi}_{m,n}(u)$$

where

$$\alpha_{m,n} = \int_0^1 F^{-1}(u)\widetilde{\psi}_{m,n}(u)du,$$

and

$$\beta_{m,n} = \int_0^1 \frac{1}{r!}\{-\log(1-u)\}^r\widetilde{\psi}_{m,n}(u)du.$$

Now, as a specific example, let us choose the orthonormal system

$$\{\widetilde{\psi}_{m,n}(u), m \geq 0, n = 0, 1, ..., 2^m - 1\},$$

for the system of Harr given by (4-1). From (4-3), we can write for $m \geq 0, n = 0, 1, ..., 2^m - 1$

$$\begin{aligned} \alpha_{m,n} &= \int_0^1 F^{-1}(u)\widetilde{\psi}_{m,n}(u)du \\ &= \int_0^1 F^{-1}(u)\psi_{m,n}(u)du \\ &= \int_0^1 F^{-1}(u)2^{m/2}\psi_{m,n}(2^m u - n)du \\ &= 2^{m/2}\int_0^1 F^{-1}(u)\left\{1_{[0,1/2)}(2^m u - n) - 1_{[1/2,1)}(2^m u - n)\right\}du \\ &= 2^{m/2}\int_0^1 F^{-1}(u)\big\{1_{[2^{-m}n, 2^{-m}n + 2^{-m-1})} \\ &\quad - 1_{[2^{-m-1}+2^{-m}n, 2^{-m}(n+1))}\big\}du \\ &= 2^{m/2}\left\{\int_{n2^{-m}}^{n2^{-m}+2^{-m-1}} F^{-1}(u)du - \int_{2^{-m-1}+n2^{-m}}^{(n+1)2^{-m}} F^{-1}(u)du\right\}. \end{aligned}$$

Similarly, by using the expression in (4-5), we get for $m \geq 0, n = 0, 1, ..., 2^m - 1$

$$\begin{aligned}
\beta_{m,n} &= \int_0^1 g(u)\widetilde{\psi}_{m,n}(u)du \\
&= \int_0^1 g(u)\psi_{m,n}(u)du \\
&= \int_0^1 \frac{1}{r!}\{-\log(1-u)\}^r 2^{m/2}\psi_{m,n}(2^m u - n)du \\
&= \frac{2^{m/2}(-1)^r}{r!}\Bigg\{ \int_{n2^{-m}}^{n2^{-m}+\frac{1}{2^{m+1}}} (\log(1-u))^r du \\
&\quad - \int_{\frac{1}{2^{m+1}}+n2^{-m}}^{(n+1)2^{-m}} (\log(1-u))^r du \Bigg\}.
\end{aligned}$$

For the purpose of illustration, we consider here the logistic distribution given by (3-5), we computed approximations and bounds for $E(U_r)$ from (4-6) for $r = 1$ up to 8 with $k = 1$ up to 7. These values are presented in Table II along with the exact values.

Table II. Approximations and bounds in (4-6) for $E(U_r)$ from the logistic distribution

r k	1	2	3	4	5	6	7	Exact Value
1	1.3333 ±0.4063	1.4979 ±0.1983	1.5740 ±0.0976	1.6102 ±0.0484	1.6278 ±0.0240	1.6364 ±0.0120	1.6407 ±0.0060	1.6449
2	1.9273 ±1.3962	2.3062 ±0.8086	2.5338 ±0.4622	2.6682 ±0.2611	2.7463 ±0.1460	2.7909 ±0.0809	2.8161 ±0.0444	2.8470
3	2.1530 ±3.1967	2.7238 ±2.0452	3.1343 ±1.2920	3.4190 ±0.8042	3.6094 ±0.49303	3.7328 ±0.2979	3.8108 ±0.1776	3.9293
4	2.2248 ±6.41866	2.9105 ±4.3017	3.4698 ±2.8750	3.9097 ±1.9049	4.2411 ±1.2470	4.4811 ±0.8052	4.6489 ±0.5126	4.9662
5	2.2440 ±12.4326	2.9817 ±8.4769	3.6331 ±5.8107	4.1946 ±3.9796	4.6602 ±2.7103	5.0307 ±1.8295	5.3143 ±1.2215	5.9836
6	2.2483 ±23.9448	3.0052 ±16.4148	3.7025 ±11.3563	4.34091 ±7.8883	4.9093 ±5.4784	5.3975 ±3.7913	5.8012 pm2.6076	6.9919
7	2.2492 ±46.2207	3.0119 ±31.7347	3.72846 ±22.0188	4.4077 ±15.3714	5.0422 ±10.7603	5.6192 ±7.5327	6.1272 ±5.2617	7.9960
8	2.2493 ±89.5443	3.0137 ±61.5065	3.7371 ±42.7110	4.4350 ±29.8628	5.1061 ±20.9613	5.7408 ±14.7389	6.3266 ±10.3644	8.9980

5 Maple Procedure(**for wavelet expansion)**

restart;
> $f := unapply(log(y) - log(1-y), y):$ Enter $F^{-1}(x)$
> $for\ r\ from\ 1\ to\ 10\ do$;
> $for\ k\ from\ 1\ to\ 7\ do$;
> $g := unapply((-log(1-y))^r/r!, y):$
> $a := array(0..k, 0..2^k - 1):$
> $for\ m\ from\ 0\ to\ k\ do$;
> $for\ n\ from\ 0\ to\ 2^m - 1\ do$;

$> a[m,n] := evalf(2^(m/2)*(int(f(y),y=n*2^(-m)..n*2^(-m)+2^(-m-1))-int(f(y),y=n*2^(-m)+2^(-m-1)..(n+1)*2^(-m))));$
$end\ do;\ end\ do:$
$> unassign('m'); unassign('n');$
$> b := array(0..k, 0..2^k-1):$
$> for\ n\ from\ 0\ to\ 2^m-1\ do;$
$> b[m,n] := evalf(2^(m/2)*(int(g(y),y=n*2^(-m)..n*2^(-m)+2^(-m-1))-int(g(y),y=n*2^(-m)+2^(-m-1)..(n+1)*2^(-m))));$
$end\ do;\ end\ do;$
$> unassign('m'); unassign('n');$
$> A := int(f(y),y=0..1)*int(g(y),y=0..1)+(sum(sum(a[m,n]*b[m,n],n=0..2^m-1),m=0..k)):$
$> unassign('m'); unassign('n');$
$> B := (evalf(int(f(y)^2,y=0..1))-((int(f(y),y=0..1))^2$
$+ sum(sum((a[m,n])^2,n=0..2^m-1),m=0..k))):$
$> unassign('m'); unassign('n');$
$> C := (evalf(int(g(y)^2,y=0..1))-((int(g(y),y=0..1))^2$
$+ sum(sum((b[m,n])^2,n=0..2^m-1),m=0..k))):$
$> R := (B*C)^(1/2):$
$> M_u := evalf(int(f(y)*g(y),y=0..1));$
$> end\ do;\ end\ do:$
$\mu_r \simeq A \pm R$

References

[**1**] Ahmadi, J. (2000). *Record values, theory and applications*, Ph. D. Dissertation, Ferdowsi University of Mashhad, Iran.

[**2**] Ahmadi, J., Arghami, N. R. (2001). Universal bounds for the difference of record values, IMC33, Mazanderan University, Iran.

[**3**] Alkemade, J. A. H. (1995). The finite wavelet transform with an application to seismic processing. Wavelets, Word scientific, 183-206.

[**4**] Arnold, B. C., Balakrishnan, N., Nagaraja, H. N. (1998). *Records*, John Wiley & Sons, New York.

[**5**] Arnold, B. C. and Balakrishnan, N. (1989). Relations, bounds and approximations for order statistics, *Lecture notes in Statistics,* Vol. **53,** New York.

[**6**] Childs, A. and Balakrishnan, N. (1997). Some approximation for moments of order statistics using MAPLE. *Comput. Statist. Data Anal.*, **38**, 341 – 347.

[**7**] Cheney, W. and Light, W. (2000). *A course in approximation theory,* Brooks/ Cole publishing company.

[**8**] David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*, John Wiley & Sons, New York.

[**9**] Gajek, L. and Okolewski, A. (2000). Inequalities for generalized order statistics from some restricted family of distributions. *Commun. Statist.- Theory Meth.* **29**, 2427–2438.

[**10**] Mallat, S. (1989). Multiresolution approximation and wavelets. *Trans. Amer. Math. Soc.* , **315,** 69-88.

[**11**] Nagaraja, H. N. (1978). On the expected values of record values. *Austral. J. Statist.*, **20**, 176 – 182.

[**12**] Raqab, M. Z. and Rychik, T. (2002). Sharp bounds for the mean of k-th record value. *Commun. Statist.- Theory Meth.* **31**, 1927–1937.

[**13**] Raqab, M. Z. (2000). On the moment of record values. *Commun. Statist.- Theory Meth.* **29**, 1631–1647.

[**14**] Walker, J. S. (1997). Fourier analysis and wavelet analysis, *Notices of the AMS*, **44**, 658-668.

Using the probability theory for matching of structural data

Alireza Ahmadyfard

Department of Electrical Eng and Robotics, Shahrood University of Technology, Shahrood, Iran

Abstrcat: The probability theory has been frequently used for pattern matching. In this context we refer to a structural data as *pattern*. A common method for matching of two pattern is the computing the probabilities of assigning the individual components in the patterns. In this approach the relation between components within patterns is ignored where as such information is vital for recognition of patterns with similar components and different structure. In this paper we propose a method in which the consistency between components in two patterns are measured in a probabilistic way. Based on this measure we estimate the probabilities of assigning the components in two patterns. This estimation is gradually improved in an iterative process.

We apply our method for matching of object images against their models. The results of experiment in real scenario confirms the superiority of this method to the direct matching and previous techniques.

Keywords: Probabilistic pattern matching, Graph matching.

1 Introduction

Matching of similar patterns is one of challenging problems in computer science and an elegant solution to this problem has many applications [Lu][Wong]. We refer to any structural data consisting of a number of components arranged in a specific configuration as a pattern. A molecule, a hand written word [Lu] or a image of specific object can be considered as a pattern [Besl]. For instance in the latter case the image of an object can be segmented to several parts where each part is geometrically related to the other parts. Figure 1 shows segmented regions of an object in its image.

In order to recognize a given pattern, it has to be compared with the category of possible patterns to identify for instance what means a hand written word or which object exists in a given image [Pope]. This problem is solved by measuring the similarity between a given pattern against all model patterns. The model with the highest similarity measure determines the identity of the test pattern.

A common solution to this problem is to compute the probability that the components in the two patterns are matched pairwise regardless of relations between the components in the patterns [Basri]. Despite of this method, the solution has several drawbacks. First of all it is likely to have several patterns with similar components but different structure. Second, the presence of similar components in a pattern results an ambiguity when components are compared pairwise. This consideration suggests us to involve the relation between components in each pattern when matching takes place.

Graphs are powerful tools for representing a structural data [Lu]. In the graph representation of a pattern each node represents a pattern component and the relation between two components in the pattern is represented by an edge in the graph which connects the corresponding nodes.

In this paper we propose a probabilistic algorithm for matching of two pattern. The probability that a node (component in pattern) in pattern A matches to a candidate node in pattern B is computed by considering not only similarity between these components, but also by measuring the consistency of their neighbouring nodes with this node assignment.

The paper is organized as follows. In the following Section we described the representation of an arbitrary pattern using graph and our algorithm for matching of patterns. The results of experiments on real patterns are given in Section 3. Section 4 draws the paper to conclusion.

2 Methodology

We represent pattern $\mathbf{Q}$ using the attributed graph $\mathbf{GQ}$ where node θ_i represent ith component in the pattern. In this regard we describe node i using vector A_{Qi}. Moreover if there is any relation between the component ith and jth in pattern $\mathbf{Q}$, an edge connects the corresponding nodes in the graph and such relation is described using binary measurement vector B_{Qij}.

In pattern matching problem we have a test pattern $\mathbf{T}$ represented using the graph $\mathbf{G_T}$ and a set of model patterns $M = \{\mathbf{M}_1, \mathbf{M}_2, \cdots, \mathbf{M}_N\}$, represented in graph set $\mathbf{G}_M = \{\mathbf{G_{M_1}}, \mathbf{G_{M_2}}, \cdots, \mathbf{G_{M_N}}\}$. The problem is to find the best model matched to the test pattern $\mathbf{T}$. A straightforward solution to this problem is to measure the similarly G_T (the test pattern graph) and each of model graphs in set M.

In this paper we propose a probabilistic method for matching of two patterns. In this method in addition to assessing the similarity of individual components in two patterns, the pairwise relations between the components are considered.

Let us assume that the test graph $\mathbf{G_T}$ has K nodes denoted by set $\Theta = \{\theta_1, \theta_2, \cdots, \theta_K\}$ and model graph $\mathbf{G_{M_m}}$ has L

nodes in set $\Omega_m = \{\omega_0, \omega_1, \omega_2, \cdots, \omega_L\}$. The problem of pattern matching can be translated as the problem of finding a correspondence between the components of the these two sets (Θ and Ω_m). The virtual node ω_0 is added to Ω_m and will be matched to any node in Θ for which none of real nodes in Ω_m is suitable. Adding this node enable us to apply the total probability theorem on this assignment.

$$\sum_{l=0}^{L} P(\theta_i \leftarrow \omega_l) = 1 \tag{1}$$

A simple node assignment is performed as follow:

$$P(\theta_i \leftarrow \omega_j) > P(\theta_i \leftarrow \omega_l) \quad \forall l \neq j \tag{2}$$

This means that ω_j will corresponds to θ_i if the corresponding probability is the highest among others. If the test graph $\mathbf{G_T}$ is dissimilar to the model graph $\mathbf{G_{M_m}}$ ideally all nodes in $\mathbf{G_T}$ must assign to the null. Thus the ratio of the number of nodes in $\mathbf{G_T}$ assigned to real nodes in $\mathbf{G_{M_m}}$ to the number of nodes assigned to null can be regarded as a measure for similarity of two patterns.

For matching of graphs many approaches have been proposed [Bunke] [Read] [Rosenfeld] [Wilson]. Among these methods some strictly compare the structure of two graphs [Bunke] [Read] where as other methods [Rosenfeld] [Wilson] tolerate more for structural difference between graphs. In fact a proper matching method must tolerate for some amount of error between two similar patterns otherwise in realistic situations the method would fail.

One of statistical techniques for matching of graphs in presence of low level errors is relaxation labeling technique [Rosenfeld]. The technique was originally inadequate for matching of real patterns. In later work, Christmas et al [Christmas] modified the original relaxation labeling technique and applied it on road matching problem. Their method works well on real data as long as the level of error between two graphs is low. Based on our preliminary experiment the reason for this failure is the product in the support function [Christmas] which drives the total support to zero.

To improve the robustness of the technique in presence of noise we have adopted the benevolent sum support function to measure the supporting evidence from the neighbouring nodes in graphs. Our matching technique is as follow.

Let us compute the probability that each node in $\mathbf{G_T}$ (the graph of test pattern) corresponds to each node in the model graph $\mathbf{G_{M_m}}$. This is the initialization stage of method. These probabilities are initialized based on the node attributes.

$$P^{(0)}(\theta_i \leftarrow \omega_j) = P(\theta_i \leftarrow \omega_j | A_{T_i}) \tag{3}$$

Applying the Bayes theorem we have:

$$P^{(0)}(\theta_i \leftarrow \omega_j) = \frac{p(A_{T_i}|\theta_i \leftarrow \omega_j)P(\theta_i \leftarrow \omega_j)}{\sum_{\omega_\alpha \in \Omega_m} p(A_{T_i}|\theta_i \leftarrow \omega_\alpha)P(\theta_i \leftarrow \omega_\alpha)} \tag{4}$$

with the normalization carried out over list of nodes in Ω_m including the null. Let ζ be the proportion of the test pattern nodes that will assume the null node. Then the prior assignment probabilities will be given as:

$$P(\theta_i \leftarrow \omega_\lambda) = \begin{cases} \zeta & \lambda = 0 \text{ (null label)} \\ \frac{1-\zeta}{L+1} & \lambda \neq 0 \end{cases} \tag{5}$$

where $L+1$ is the number of nodes in the model graph G_{M_m}.
We shall assume that the errors on node attributes are statistically independent and their distribution function is Gaussian, i.e.

$$p(A_{T_i}|\theta_i \leftarrow \omega_\alpha) = N_{A_{T_i}}(A_{m_\alpha}, \Sigma_u) \tag{6}$$

where Σ_u is a diagonal covariance matrix for attributed vector A_{T_i} corresponding to node θ_i in the test pattern graph. The vector A_{m_α} is the attribute vector for node ω_α in the model graph.

As mentioned earlier the goal of our method is to involve relation between components in two patterns under match. Thus we use pairwise information between components in the test and model graphs to modify the node matching probabilities computed at the initialization stage. This process is performed in an iterative manner. We consider all possible matching for node θ_i and the related probabilities using their previous values and supports provided by the nodes connected to the θ_i.

$$P^{(n+1)}(\theta_i \leftarrow \omega_j) = \frac{P^{(n)}(\theta_i \leftarrow \omega_j)Q^{(n)}(\theta_i \leftarrow \omega_j)}{\sum_{\omega_\lambda \in \Omega_m} P^{(n)}(\theta_i \leftarrow \omega_j)P^{(n)}(\theta_i \leftarrow \omega_\lambda)} \tag{7}$$

$$\begin{aligned} Q^{(n)}(\theta_i \leftarrow \omega_\alpha) = & P^{(n)}(\theta_i \leftarrow \omega_\alpha) \sum_{\theta_l \in \Theta} \sum_{\omega_\beta \in \Omega_m} P^{(n)}(\theta_l \leftarrow \omega_\beta) \\ & \times p(B_{il}|\theta_i \leftarrow \omega_\alpha, \theta_j \leftarrow \omega_\beta) \end{aligned} \tag{8}$$

where function Q quantifies the support that assignment $(\theta_i \leftarrow \omega_\alpha)$ receives at the nth iteration step from the connected nodes to the node θ_i in the test pattern.

In the support function, Q, the term $p(B_{il}|\theta_i \leftarrow \omega_\alpha, \theta_l \leftarrow \omega_\beta)$ behaves as a compatibility coefficient in other relaxation methods [Rosenfeld]. In fact it is the density function for the binary measurement vector B_{il} given the matches $\theta_i \leftarrow \omega_\alpha$ and $\theta_l \leftarrow \omega_\beta$. Similarly for the distribution function of binary relations is centered on the model binary vector $B_{m_{\alpha\beta}}$. It is assume that deviations from this mean are modeled by a Gaussian. Thus we have:

$$p(B_{il}|\theta_i \leftarrow \omega_\alpha, \theta_l \leftarrow \omega_\beta) = N_{B_{il}}(B_{m_{\alpha\beta}}, \Sigma_b) \tag{9}$$

where Σ_b is the covariance matrix of the binary vector B_{il}.
The iterative proccess will be terminated in one of the following circumstances:

1. If in the last iteration none of the probabilities changed by more than threshold ϵ.

2. If the number of iterations reached some specified limit

After terminating the proccess for each node θ_i only one node from Ω_m is assigned. The corresponding probability would be very close to one and the probabilities of other node assignments would be close two zero. As a proper measure for the similarity between the test and a model pattern. We compute the ratio of the number of nodes in the test pattern assigned to non-null to the total number of nodes in the pattern. The model with the highest similarity measure determines the identity of the test pattern.

3 Experimental Results

In this section we present the results of experiments designed to demonstrate the advantages of the proposed method of probabilistic pattern matching.

We apply our method for recognition of objects from their **2D** images. For this experiment we used the frontal view of objects as their models. The 20 images of objects taken from -80 to +80 degrees are used as the test images. We used 100 objects in the COIL database for test. So we have 100 model pattern and 100×20 test patterns. Each image is first segmented to its constituent regions considered as the pattern components of the object image (pattern). Between each component (a region) and each of its neighbouring regions a connection is provided as an edge in the graph (Figure 1).
As the attribute vector of each component in an object image we use the region colour in RGB (Red Green Blue) space. Associated to any two neighbouring regions (i,1) (the connected nodes in the pattern graph) we define a binary measurement B_{il} using the ratio of the area of two regions. This binary measurement in general can be a vector of several measures. We applied our method on the introduced data, after 5 iterations the algorithm converges so that the assignment probabilities are either close to one or zero. The result of image matching has been shown in figure 2. We also apply the Christmas's method and the direct matching (matching

of individual components regardless of neighbourhood information). The measure of assessment is the correct assignment of the models to the test images (finding correctly the identity of objects in the test images).

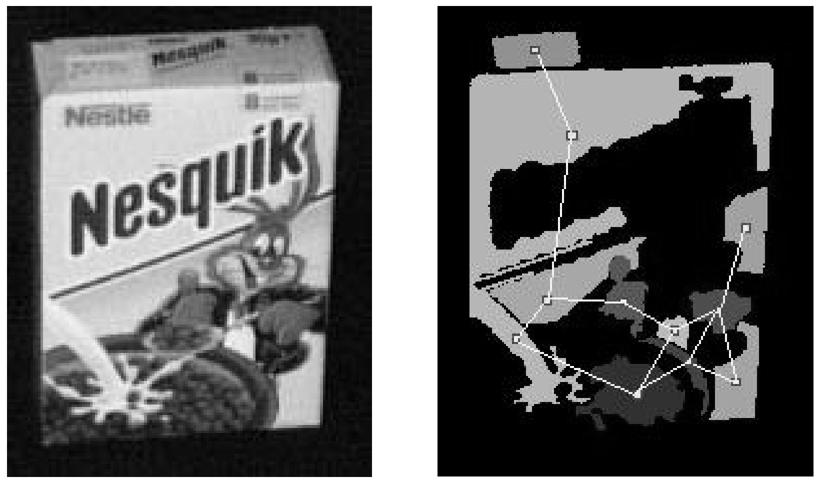

Fig. 1. a) an image of the object b) the image components represented in a graph

In this figure we compared the results of image matching using our method with Christmas et al method [Christmas] and the direct matching. We need to mention that in direct matching the relation between components in a pattern is ignored. Thus he matching of individual components (regions) regardless of their relations is performed. As shown in this figure both our method and Christmas et al [Christmas] method works better than the direct matching for all 20 views of objects. The figure also shows that our method also performs better than Christmas's method for the most of object views. In all three methods as the figure shows the maximum recognition rate occurs at the view #20 because this view of an objects is the same as the frontal view which is used as the model of the object.

4 Conclusions

We addressed the problem of probabilistic pattern matching which has many application in computer science. We represented the model and test patterns in the form of graphs. A new probabilistic relaxation technique was proposed. In this method the probabilities of node assignment were modified in a iterative process using the pairwise relation between the connect nodes in the model and test patterns. The experimental results obtained on real data demonstrated that the proposed method can deliver a pattern matching performance better than the direct matching and the previous method.

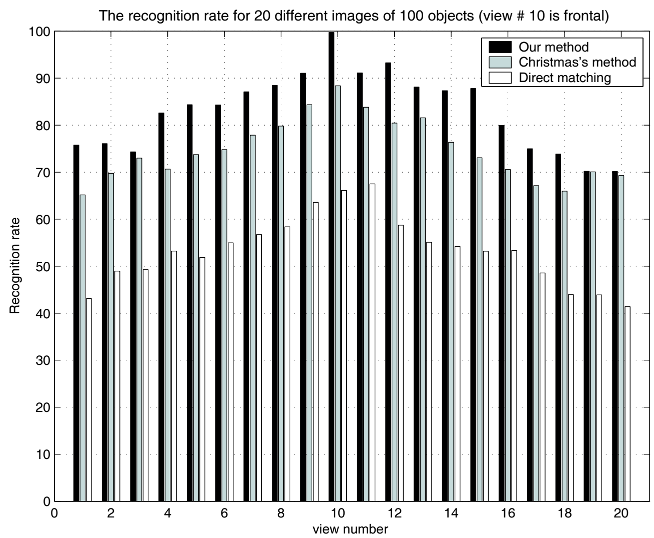

Fig. 2. The comparison between the recognition rates in the three metods. This rates are plotted against object views (0 to 20) which are corresponds to -80 to 80 degrees

References

[**1**] Basri, R. and Gacobs, D. (2000), Projective alignment with regions. Technical Report Technical Report, MCS00-19, Mathematics and Computer Science, Weizmann Institute of Science.

[**2**] Besl, P. and Jain, R.J. (1985), Three-dimensional object recognition. ACM Computing Survey, 17, 75-145.

[**3**] Bunke, H. and Shearer, K. (1998), A graph distance metric based on the maximal common subgraph. Pattern Recognition Letters, 255-259.

[**4**] Christmas, W., Kittler, J. and Petrou, M. (1995), Structural matching on computer vision using probabilistic relaxation. IEEE Transaction on Pattern Analysis and Machine Intelligence, 749-764.

[**5**] Lu, S.W., Ren, Y. and Suen, C.Y. (1991), Hierarchical Attributed Graph representation and recognition of handwritten Chinese characters. Pattern Recognition, 24, 617-632.

[**6**] Pope, A.R. (1994), Model-based object recognition, a survey of recent research. Technical report, University of Columbia.

[**7**] Read, R.C. and Corneil, D.G. (1977), The graph isomorphism disease. Graph Theory, 1.

[**8**] Rosenfeld, A., Hummel, R. and Zucker, S. (1976), Scene labeling by relaxation operations. IEEE Transaction on system man and cybernetics, 420-433.

[**9**] Wilson, R., Cross, A. and Hancock, E. (1997), Inexact graph matching with genetic search. Pattern Recognition, 953-970.

[**10**] Wong, A. and You, M. (1985), Entropy and distance of random graphs with application to structural pattern recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 599-609.

On the average number of sharp crossings of certain gaussian random polynomials

S. Rezakhah, S. Shemehsavar

Department of Mathematics Amirkabir University of Technology Hafez Avenue Tehran, Iran

Abstrcat: We study a random algebraic polynomial $Q_n(x) = \sum_{i=0}^{n} A_i x^i$ where the coefficients $A_0, A_1, \cdots$ form a sequence of centered Gausian random variables. Moreover, assume that the increments $\Delta_j = A_j - A_{j-1}$, $j = 0, 1, 2, \cdots$ are independent, $A_{-1} = 0$. The coefficients can be considered as n consecutive observations of a Brownian Motion. We obtain the asymptotic behaviour of the expected number of sharp crossings, with slope $u = O(n)$, of polynomial $Q_n(x)$. Sharp crossings are those zero up-crossing, with slope greater than u, or those down-crossing with slope smaller than $-u$. We consider the cases where u is unbounded and is increasing with n.

Keywords: random algebraic polynomial, number of real zeros,sharp crossing, expected density, Brownian Motion

1 Preliminaries

In this paper we study the expected number of sharp crossings of the random algebraic polynomial $Q_n(x) = \sum_{i=1}^{n} A_i x^i$, where the coefficients A_i are consecutive observations of a Brownian motin. Let $S_u(a,b)$, be the number of u-sharp zero crossing of $Q_n(x)$ in the interval (a,b), i.e. those zeros up-crossing with slope greater than u or down-crossing with slope smaller than $-u$. We obtain the asymptotic behavior of the expected number of $S_u(-\infty, \infty)$. It is said that $Q_n(x)$ have a zero up-crossing at t_0 if there exists $\varepsilon > 0$ such that $Q_n(x) \leq 0$ in $(t_0 - \varepsilon, t_0)$ and $Q_n(x) > 0$ in $(t_0, t_0 + \varepsilon)$, or zero down-crossing at t_0 if the above inequalities have reverse sides.

Farahmand [3] considered the case where the coefficients A_i are independent normally distributed and obtained the asymptotic behavior of the expected number of u-sharp crossings of $Q_n(x)$. Here we consider a more general case where the coefficients are not independent nor identically distributed.

Definition 1.1: The expected number of u-sharp crossing of a process $Q_n(x)$ is defined by

$$ES_u(a,b) = \int_a^b dt \{ \int_{-\infty}^{-u} + \int_u^{\infty} \} |y| p_t(0,y) dy$$

where $p_t(0, y)$ denotes the joint density of $Q_n(x)$ and its derivative $Q'_n(x)$.

We study the asymptotic behavior of the expected number of u-sharp crossings of $Q_n(x)$, for any positive $u = O(n)$.

The theory of the expected number of real zeros of random algebraic polynomials was elaborated by the fundamental work of M. Kac [4]. The work of Edelman and Kostlan [2] provides a nice approach to the subject, and also gives main references up to 1995. There are recent interests in cases that the coefficients form certain random processes, Rezakhah and Soltani [5,6].

2 Expected number of sharp-crossings

Let $A_0, A_1, \cdots$ be a mean zero Gaussian random sequence for which the increments $\Delta_i = A_i - A_{i-1},\ i = 1, 2, \cdots$ are independent, $A_{-1} = 0$. The sequence $A_0, A_1 \cdots$ may be considered as successive Brownian points, i.e., $A_j = W(t_j),\ j = 0, 1, \cdots$, where $t_0 < t_1 < \cdots$ and $\{W(t),\ t \geq 0\}$ is the standard Brownian motion. In this physical interpretation, $\mathrm{Var}(\Delta_j) = t_j - t_{j-1}$, the distance between successive times.

Let

$$Q_n(x) = \sum_{i=0}^{n} A_i x^i, \qquad -\infty < x < \infty, \tag{2.1}$$

where A_j's are defined as above, and $S_u(a, b)$ denote the number of u-sharp zero crossings of $Q_n(x)$ in the interval (a, b). We note that $A_j = \Delta_0 + \Delta_1 + \cdots + \Delta_j,\quad j = 0, 1, \cdots$, where $\Delta_i \sim N(0, \sigma_i^2)$ and Δ_i are independent, $i = 0, 1, \cdots$. Thus

$$Q_n(x) = \sum_{k=0}^{n} (\sum_{j=k}^{n} x^j) \Delta_k = \sum_{k=0}^{n} a_k(x) \Delta_k,$$

and

$$Q'_n(x) = \sum_{k=0}^{n} (\sum_{j=k}^{n} j x^{j-1}) \Delta_k = \sum_{k=0}^{n} b_k(x) \Delta_k,$$

where

$$a_k(x) = \sum_{j=k}^{n} x^j, \quad b_k(x) = \sum_{j=k}^{n} j x^{j-1}, \qquad k = 0, \cdots, n. \tag{2.2}$$

Let $ES_u(a,b)$ denote the expected number of u-sharp zero crossings of $Q_n(x) = 0$ in any interval (a,b). The classical result of Cramer and Leadbetter [1967 p 285] can be applied to obtain;

$$ES_u(a,b) = \int_a^b f_n(x)dx \tag{2.3}$$

where

$$f_n(x) = \frac{1}{\pi} g_{1,n}(x) \exp(g_{2,n}(x) \qquad (2.4))$$

and

$$g_{1,n}(x) = EA^{-2}, \;\; g_{2,n}(x) = -\frac{A^2u^2}{2E^2}, \tag{2.5}$$

where

$$A^2 = \mathrm{Var}(Q_n(x)) \;=\; \sum_{k=0}^{n} a_k^2(x)\sigma_k^2, \quad B^2 = \mathrm{Var}(Q'_n(x)) = \sum_{k=0}^{n} b_k^2(x)\sigma_k^2,$$

$$D = \mathrm{Cov}(Q_n(x), Q'_n(x)) = \sum_{k=0}^{n} a_k(x)b_k(x)\sigma_k^2, \quad \text{and} \quad E^2 = A^2B^2 - D^2,$$

and $a_k(x)$, $\quad b_k(x)$ is defined by (2.2).

3 Asymptotic behaviour of ES_U

If $\Delta_1 \cdots \Delta_n$ have the same distributions, then σ_k^2 can be taken to be one, i.e., $\sigma_k^2 = 1$, $k = 1 \cdots n$. We also assume that $u = O(n)$.
Theoream 3.1: Let $Q_n(x)$ be the random algebric polynomial given by (2.1), for which $A_j = \Delta_1 + ... + \Delta_j$ where $\Delta_j \sim N(0,1)$, and are independent for $j = 1, ..., n$. Then the expected number of u-sharp zero crossings of $Q_n(x) = 0$, for large n satisfies:

$$\begin{aligned} ES_u(-\infty,\infty) \;\sim\; & \frac{1}{\pi\sqrt{2n}}\left(-\pi + 2\arctan(\frac{1}{2\sqrt{2n}})\right) + \frac{1}{\pi}\log(2n+1) \\ & + \frac{1}{\pi}(1.920134502) + \frac{C1}{n\pi} \\ & + \frac{u^2}{n^3\pi}\left(20.53664050 - 2.66666667\ln(n^3+1)\right) + O(\frac{1}{n^2}) \end{aligned}$$

where $C1 = -.7190843666$ for n even and $C1 = 1.716159419$ for n odd.
proof: Due to the behaviour of $Q_n(x)$, the asymptotic behaviour is treated separately on the intervals $1 < x < \infty$, $-\infty < x < -1$, $0 < x < 1$ and

$-1 < x < 0$. For $1 < x < \infty$, by the change of variable $x = 1 + \frac{t}{n}$ and the equality $\left(1+\frac{t}{n}\right)^n = e^t\left(1-\frac{t^2}{n}\right) + O\left(\frac{1}{n^2}\right)$. Using (2.3), we find that

$$ES_u(1,\infty) = \frac{1}{n}\int_0^\infty f_n\big(1+\frac{t}{n}\big)dt,$$

where by (2.5) and by tedious manipulation we have that

$$g_{2,n}\left(1+\frac{t}{n}\right) = o(n^{-2})$$

and

$$n^{-1}g_{1,n}\left(1+\frac{t}{n}\right) = \left(R_1(t) + \frac{S_1(t)}{n} + O\left(\frac{1}{n^2}\right)\right), \quad n\to\infty, \qquad (3.1)$$

where

$$R_1(t) = \frac{\sqrt{(4t-15)e^{4\,t}+(24t+32)e^{3\,t}-e^{2\,t}(8t^3+12t^2+36t+18)+8e^tt+1}}{2\,t\,(-1-3\,e^{2\,t}+4\,e^t+2\,te^{2\,t})}$$

and

$$\begin{aligned} S_1(t) \;=\; & -4\left(\left(\frac{21}{8}\,t+1/2\,t^4-2\,t^2+1/2\,t^3-\frac{153}{16}\right)e^{2\,t}\right.\\ & +\left(6\,t^3+\frac{55}{4}\,t+t^5-\frac{331}{16}-\frac{53}{4}\,t^2-9/2\,t^4\right)e^{4\,t}\\ & +\left(\frac{41}{2}-13/2\,t^3+8\,t^2-21/2\,t\right)e^{3\,t}\\ & +\left(\frac{39}{4}+\frac{29}{4}\,t^2-3/2\,t^3-\frac{21}{4}\,t\right)e^{5\,t}+\left(-3/8\,t+1/4\,t^2-\frac{27}{16}\right)e^{6\,t}\\ & \left.+\left(7/4-1/4\,t-1/4\,t^2\right)e^t-1/16\right)\left(2e^{2t}t-1+4\,e^t-3\,e^{2\,t}\right)^{-2}\\ & \times(1+8e^tt-(18+36t+12t^2+8t^3)e^{2t}+(4t-15)e^{4t}+(32+24t)e^{3t})^{-1/2}\end{aligned}$$

One can easily verify that as $t\to\infty$,

$$R_1(t) = \frac{1}{2t^{3/2}} + O(t^{-5/2}), \qquad\qquad S_1(t) = -\frac{1}{8t^{1/2}} + O(t^{-3/2})$$

As (3.1) can not be integrated term by term, by noting that

$$\frac{I_{[t>1]}}{8n\sqrt{t}} = \frac{I_{[t>1]}}{8n\sqrt{t}+t\sqrt{t}} + O\left(\frac{1}{n^2}\right) \qquad (3.2)$$

where

$$I_{[t>1]} = \begin{cases} 1 & \text{if} \quad t\geq 1 \\ 0 & \text{if} \quad t<1 \end{cases}$$

Thus by (3.2) we have that

$$\frac{1}{n} f_n\left(1+\frac{t}{n}\right)+\frac{I_{[t>1]}}{\pi(8n\sqrt{t}+t\sqrt{t})}=\frac{R_1(t)}{\pi}+\frac{1}{\pi}\left(\frac{S_1(t)}{n}+\frac{I_{[t>1]}}{8n\sqrt{t}}\right)+O\left(\frac{1}{n^2}\right).$$

This expression is term by term integrable, and provides that

$$\begin{aligned} ES_u(1,\infty) &= \frac{1}{n}\int_0^\infty f_n\left(1+\frac{t}{n}\right)dt = \frac{1}{2\pi\sqrt{2n}}\left(-\pi+2\arctan\left(\frac{1}{2\sqrt{2n}}\right)\right) \\ &\quad +\frac{1}{\pi}\int_0^\infty R_1(t)dt+\frac{1}{\pi n}\int_0^\infty\left(S_1(t)+\frac{I_{[t>1]}}{8\sqrt{t}}\right)dt \\ &\quad +O\left(\frac{1}{n^2}\right) \end{aligned}$$

where $\int_0^\infty R_1(t)dt = .7348742023$, and $\int_0^\infty \left(S_1(t)+\frac{I_{[t>1]}}{8\sqrt{t}}\right)dt = -.2496371198.$

For $-\infty < x < -1$, let $x=-1-\frac{t}{n}$, to obtain

$$ES_u(-\infty,-1)=\frac{1}{n}\int_0^\infty f_n\left(-1-\frac{t}{n}\right)dt.$$

Using (2.5) we have that

$$g_{2,n}\left(-1-\frac{t}{n}\right)=\left(\frac{u^2}{n^3}\right)g_{2,1}(t)+O\left(\frac{1}{n^2}\right)$$

where

$$g_{2,1}(t)=\frac{16\left(2\,e^{2\,t}t+e^{2\,t}-1\right)t^3}{2\,e^{2\,t}-e^{4\,t}+12\,e^{2\,t}t^2-1+8\,e^{2\,t}t^3-4\,e^{4\,t}t+4\,e^{2\,t}t}$$

and

$$n^{-1}g_{1,n}\left(-1-\frac{t}{n}\right)=\left(R_2(t)+\frac{S_2(t)}{n}+O\left(\frac{1}{n^2}\right)\right) \tag{3.3}$$

where

$$R_2(t)=1/2\sqrt{\frac{-2\,e^{2\,t}+e^{4\,t}+1-12\,e^{2\,t}t^2-8\,t^3e^{2\,t}+4\,te^{4\,t}-4\,te^{2\,t}}{t^2\left(e^{2\,t}-1+2\,te^{2\,t}\right)^2}}$$

Finally for n even $S_2(t)=S_{2e}(t)$, and for n odd $S_2(t)=S_{2o}(t)$, where

$$S_{2e}(t)=\frac{1}{4}\left(\frac{S_{21}(t)+S_{22}(t)}{S_{23}(t)}\right) \qquad S_{2o}(t)=\frac{1}{4}\left(\frac{S_{21}(t)-S_{22}(t)}{S_{23}(t)}\right)$$

and

$$\begin{aligned} S_{21}(t) &= 1+\left(-8\,t^4+30\,t-8\,t^3+48\,t^2-3\right)e^{2\,t} \\ &\quad +\left(3-12t+52t^2+96t^3+40t^4-16\,t^5\right)e^{4\,t}-\left(18t+4t^2+1\right)e^{6\,t} \\ S_{22}(t) &= \left(8\,t+32\,t^3+40\,t^2\right)e^{3\,t}+\left(-8\,t^2-12\,t\right)e^{5\,t}+4\,e^t t \\ S_{23}(t) &= \left(e^{4\,t}(4t+1)-2e^{2t}(1+2t+6t^2+4t^3)+1\right)^{1/2}\left(e^{2\,t}(2t+1)-1\right)^2 \end{aligned}$$

It can be seen that as $n \to \infty$,

$$R_2(t) = \frac{1}{2t^{3/2}} + O(t^{-2}) \qquad\qquad S_2(t) = \frac{-1}{8t^{1/2}} + O(t^{-3/2}).$$

$$R_2(t)g_{2,1}(t) = O\left(t^{3/2}exp(-2t)\right)$$

Now by using the equality (3.2), (3.3) can be written as

$$\begin{aligned}\frac{1}{n}f_n(-1-\frac{t}{n}) + \frac{I_{[t>1]}}{\pi(8n\sqrt{t}+t\sqrt{t})} &= \frac{1}{\pi}\left(R_2(t) + \frac{1}{n}(S_2(t) + \frac{I_{[t>1]}}{8\sqrt{t}})\right.\\ &\qquad \left.+\frac{u^2}{n^3}R_2(t)g_{2,1}(t)\right) + O\left(\frac{1}{n^2}\right)\end{aligned}$$

giving that

$$\begin{aligned}ES_u(-\infty,-1) &= \frac{1}{n}\int_0^\infty f_n(-1-\frac{t}{n}) = \frac{1}{2\pi\sqrt{2n}}\left(-\pi + 2\arctan(\frac{1}{2\sqrt{2n}})\right)\\ &\quad +\frac{1}{\pi}\int_0^\infty R_2(t)dt + \frac{1}{n\pi}\int_0^\infty\left(S_2(t) + \frac{I_{[t>1]}}{8\sqrt{t}}\right)dt\\ &\quad +\frac{u^2}{n^3\pi}\int_0^\infty R_2(t)g_{2,1}(t)dt + O\left(\frac{1}{n^2}\right)\end{aligned}$$

where $\int_0^\infty R_2(t)dt = 1.095640061$,and $\int_0^\infty R_2(t)g_{2,1}(t)dt = -.9399856028$,and for n odd

$$\int_0^\infty\left(S_2(t) + \frac{I_{[t>1]}}{8\sqrt{t}}\right)dt = -.0322863105$$

and for n even

$$\int_0^\infty\left(S_2(t) + \frac{I_{[t>1]}}{8\sqrt{t}}\right)dt = -.4677136959$$

For $0 < x < 1$, let $x = 1 - \frac{t}{n+t}$, then

$$ES_u(0,1) = \left(\frac{n}{(n+t)^2}\right)\int_0^\infty f_n\left(1-\frac{t}{n+t}\right)dt,$$

where by (1.3) we have that

$$g_{2,n}\left(1-\frac{t}{n+t}\right) = o(n^{-2}),$$

and

$$\begin{aligned}&\left(\frac{n}{(n+t)^2}\right)g_{1,n}\left(1-\frac{t}{n+t}\right)\\ &= \left(1-\frac{2t}{n}+O\left(\frac{1}{n^2}\right)\right)\left(R_3(t)+\frac{S_3(t)}{n}+O\left(\frac{1}{n^2}\right)\right)\\ &= \left(R_3(t)+\frac{S_3(t)-2tR_3(t)}{n}\right)+O\left(\frac{1}{n^2}\right). \qquad (3.4)\end{aligned}$$

We observe that $R_3(t) \equiv R_1(-t)$ and

$$\begin{aligned}
S_3(t) \;=\; & -7\left(\left(-2/7\,t^4-\frac{16}{7}\,t^2-5/2\,t+6/7\,t^3-\frac{123}{28}\right)e^{-2\,t}\right.\\
&+\left(\frac{12}{7}\,t^3+4/7\,t^5-\frac{22}{7}\,t^4-13\,t^2-7\,t+9/4\right)e^{-4\,t}\\
&+\left(\frac{9}{7}+1/7\,t^2+1/7\,t\right)e^{-t}+\left(\frac{30}{7}+\frac{44}{7}\,t^2+6/7\,t^3+\frac{66}{7}\,t\right)e^{-3\,t}\\
&\left.+\left(-\frac{39}{7}+\frac{55}{7}\,t^2-6/7\,t^3-5\,t\right)e^{-5\,t}+\left(9/4+t^2+\frac{69}{14}\,t\right)e^{-6\,t}-\frac{3}{28}\right)\\
&\times\left(-3\,e^{-2\,t}+4\,e^{-t}-1-2\,e^{-2\,t}t\right)^{-2}\\
&\times\left(1-8e^{-t}t+(-18+36t-12t^2+8t^3)e^{-2t}+(-4t-15)e^{-4t}\right.\\
&\left.\qquad+(32-24t)e^{-3t}\right)^{-1/2}.
\end{aligned}$$

One can easily verify that $R_3(t)=\frac{1}{2t}+O(e^{-t/2})$, and $S_3(t)=\frac{3}{4}+O(e^{-t})$. Now by using the fact that

$$\frac{I_{[t>1]}}{2t}-\frac{I_{[t>1]}}{4n+2t}=\frac{I_{[t>1]}}{2t}-\frac{I_{[t>1]}}{4n}+O\left(\frac{1}{n^2}\right) \tag{3.5}$$

one can write (2.4) as

$$\begin{aligned}
\frac{n}{(n+t)^2}f_n(1-\frac{t}{n+t}) \;=\; & \frac{1}{\pi}\left(R_3(t)-\frac{I_{[t>1]}}{2t}\right)+\frac{1}{\pi}\left(\frac{I_{[t>1]}}{2t}-\frac{I_{[t>1]}}{4n+2t}\right)\\
&+\frac{1}{n\pi}\left(S_3(t)-2tR_3(t)+\frac{I_{[t>1]}}{4}\right)+O\left(\frac{1}{n^2}\right).
\end{aligned}$$

Thus we find that

$$\begin{aligned}
ES_u(0,1) &= \frac{n}{(n+t)^2}\int_0^\infty f_n(1-\frac{t}{n+t})dt\\
&= \frac{1}{\pi}\int_0^\infty\left(R_3(t)-\frac{I_{[t>1]}}{2t}\right)dt+\frac{1}{2\pi}(\log(2n+1)\\
&+\frac{1}{n\pi}\int_0^\infty\left(S_3(t)-2tR_3(t)+\frac{I_{[t>1]}}{4}\right)dt+O\left(\frac{1}{n^2}\right)
\end{aligned}$$

where $\int_0^\infty\left(R_3(t)-\frac{I_{[t>1]}}{2t}\right)dt=-.2897712456,$ and

$$\int_0^\infty\left(S_3(t)-2tR_3(t)+\frac{I_{[t>1]}}{4}\right)dt=.498174649.$$

For $-1<x<0$, using the change of variable $x=-1+\frac{t}{n+t}$, we have that $ES_u(-1,0)=\left(\frac{n}{(n+t)^2}\right)\int_0^\infty f_n\left(-1+\frac{t}{n+t}\right)dt,$ where by (2.5)

$$g_{2,n}\left(-1+\frac{t}{n+t}\right)=\left(\frac{u^2}{n^3}\right)g_{2,1}(t)+O(\frac{1}{n^2}),$$

in which

$$g_{2,1}(t) = -\frac{16\left(-e^{-2\,t}+1+2\,e^{-2\,t}t\right)t^3}{\left(-4\,e^{-4\,t}t+e^{-4\,t}-12\,e^{-2\,t}t^2+4\,e^{-2\,t}t+1+8\,e^{-2\,t}t^3-2\,e^{-2\,t}\right)},$$

and

$$\left(\frac{n}{(n+t)^2}\right)g_{1,n}(t) = \left(\frac{n^2}{(n+t)^2}\right)\left(R_4(t)+\frac{S_4(t)}{n}+O\left(\frac{1}{n^2}\right)\right), \qquad (3.6)$$

where $R_4(t) \equiv R_2(-t)$. For n even $S_4(t) \equiv S_{4e}(t) = \frac{1}{4}\left(\frac{S_{41}(t)+S_{42}(t)}{S_{43}(t)}\right)$, and for n odd $S_4(t) \equiv S_{4o}(t) = \frac{1}{4}\left(\frac{S_{41}(t)-S_{42}(t)}{S_{43}(t)}\right)$, where

$$\begin{aligned} S_{41}(t) \;=\; & 8\left(-9/4\,t+6\,t^2-3\,t^3-\frac{9}{8}+t^4\right)e^{-2\,t} \\ & +8\left(15\,t^4-3/2\,t-22\,t^3+\frac{9}{8}+19/2\,t^2-2\,t^5\right)e^{-4\,t} \\ & +8\left(\frac{15}{4}\,t-7/2\,t^2-3/8\right)e^{-6\,t}+3, \end{aligned}$$

$S_{42}(t) \equiv S_{22}(-t)$, and $S_{43}(t) \equiv S_{23}(-t)$. As $t \to \infty$ we have

$$R_4(t) = \frac{1}{2t}+o(e^{-t}), \quad S_4(t) = \frac{3}{4}+O(te^{-t}), \quad R_4(t)g_{2,1}(t) = -8t^2+O(t^2e^{-2t})$$

Using the equality $\frac{n^2}{(n+t)^2} = 1-2\frac{t}{n}+O\left(\frac{1}{n^2}\right)$, we have that

$$\begin{aligned} \frac{n^2}{(n+t)^2}f_n(-1+\frac{t}{n+t}) \;&=\; \frac{1}{\pi}\left(1-\frac{2t}{n}+O(\frac{1}{n^2})\right)\left(R_4(t)+\frac{S_4(t)}{n}+O(\frac{1}{n^2})\right) \\ & \quad\times\left(1+\frac{u^2}{n^3}g_{2,1}(t)+O(\frac{1}{n^2})\right) \end{aligned}$$

This by (3.5) and (2.4) leads to

$$\begin{aligned} ES_u(-1,0) &= \frac{n^2}{(n+t)^2}\int_0^\infty f_n(-1+\frac{t}{n+t})dt \\ &= \frac{1}{2\pi}(\log(2n+1))+\frac{1}{\pi}\Bigg\{\int_0^\infty(R_4(t)-\frac{I_{[t>1]}}{2t})dt \\ &+\frac{1}{n}\int_0^\infty(S_4(t)-2tR_4(t)+\frac{I_{[t>1]}}{4})dt \\ &+\frac{u^2}{n^3}\int_0^\infty(R_4(t)g_{2,1}(t)+8t^2)dt-u^2\int_0^\infty\frac{8t^2}{n^3+exp(t^3)}dt\Bigg\} \\ &+O(\frac{1}{n^2}) \end{aligned}$$

where for n even

$$\int_0^\infty(S_4(t)-2tR_4(t)+\frac{I_{[t>1]}}{4})dt = -.4999081999$$

and for n odd

$$\int_0^\infty (S_4(t) - 2tR_4(t) + \frac{I_{[t>1]}}{4})dt = 1.499908200.$$

Finally we have that

$$\int_0^\infty (R_4(t) - \frac{I_{[t>1]}}{2t})dt = .3793914851, \qquad \int_0^\infty (R_4(t)g_{2,1}(t) + 8t^2)dt = 21.47662610,$$

and

$$\int_0^\infty \frac{8t^2}{n^3 + exp(t^3)}dt = 2.666666667ln(n^3+1)/n^3,$$

which leads to the result of theorem (3.1).

References

[**1**] Cramér, H and Leadbetter, M. R. (1967). *Stationary and Related Stochastic Processes*, John Wiley N.Y.

[**2**] Edelman, A. and Kostlan, E. (1995). How many zeros of a random polynomial are real? Bull. Amer. Math. Soc., 32, 1-37.

[**3**] Farahmand, K. (1998). *Topics in Random Polynomials*, Chapman & Hall

[**4**] Kac, M. (1943). On the average number of real roots of a random algebraic equation, Bull. Amer. Math. Soc., 49, 314-320.

[**5**] Rezakhah, S. and Soltani, A. R. (2000). On the Levy and Harmonizable random polynomials. Georgian. Math. Journal, vol. 7, No. 2, pp. 379-386.

[**6**] Rezakhah, S. and Soltani, A. R. (2003). On the Expected Number of Real Zeros of Certain Gaussian Random Polynomials. Stoc. Anal, Appl, vol. 21, No. 1, pp. 223-234.