

**PROPERTIES OF STATISTICS CONNECTED WITH
MINIMAL SPACING
AND RECORD EXCEEDANCES**

I. BAIRAMOV AND S. ERYILMAZ

*Ankara University, Faculty of Sciences, Department of Statistics
06100, Tandoğan, Ankara, Turkey*

SUMMARY

An additional insight to the distributional properties of statistics based on minimal spacing and record exceedances is provided. Behaviour of a sequence of independent identically distributed random variables with respect to a random threshold constructed by the help of minimal spacing and records is studied.

1. INTRODUCTION

This paper attempts to provide an additional insight into properties of exceedances and placement statistics described in Bairamov and Eryılmaz (2000). The importance of these concepts in construction of non-parametric tests of equality (identity) of distributions has been demonstrated by Katzenbeisser (1985), (1986), Matveychuk and Petunin (1990), Johnson and Kotz (1991), (1994), Wesolowsky and Ahsanullah (1998), Bairamov, Gebizlioglu and Kaya (1999). Our study is also related to the theory of invariant confidence intervals which contain the main distributed mass or future observations. Invariant confidence intervals which involve future observations are similar to, but not identical with, the approach of tolerance limits. Tolerance limits are originally introduced by Shewart (1931). Tolerance limits and exceedance statistics are widely discussed by Wilks (1941), Robbins (1944), Gumbel and Shelling (1950), Epstein (1954), Sarkadi (1957), Siddiqui (1977) along with others. Further detailed discussion can be found in

David (1981), Leadbetter at all (1983), Johnson at all (1992). A recent interest in the subject lies in the work of Wesolowsky and Ahsanullah (1998), Bairamov and Kotz (2000).

Invariant confidence intervals containing the main distributed mass were introduced in 1990 (c.f., Bairamov and Petunin (1991)). We mention here the definition and some results concerning invariant intervals. Let X_1, X_2, \dots, X_n be a random sample of size n with distribution function (d.f.) $F \in \mathfrak{F}$, \mathfrak{F} is some class of d.f.'s. $f_1(\cdot)$ and $f_2(\cdot)$ are assumed to be two Borel functions satisfying $f_1(x_1, x_2, \dots, x_n) \leq f_2(x_1, x_2, \dots, x_n), \forall (x_1, x_2, \dots, x_n) \in R^n$. The random interval $(f_1(X_1, X_2, \dots, X_n), f_2(X_1, X_2, \dots, X_n))$ is called an invariant confidence interval containing the main distributed mass (or invariant confidence interval containing the future observations) for class \mathfrak{F} , if $\exists \beta \in (0, 1)$ such that $P\{X_{n+1} \in (f_1(X_1, X_2, \dots, X_n), f_2(X_1, X_2, \dots, X_n))\} = \beta, \forall F \in \mathfrak{F}$. The quantity β is the same for all $F \in \mathfrak{F}$ and is called a confidence level of an invariant interval. Let $\mathfrak{F} = \mathfrak{F}_c$ be the class of all continuous d.f.'s. It is known that under some natural conditions any invariant confidence interval for the class \mathfrak{F}_c can be generated only by the order statistics, so that $f_1(X_1, X_2, \dots, X_n) = X_{(i)}, f_2(X_1, X_2, \dots, X_n) = X_{(j)}, 1 \leq i < j \leq n$, where $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ are order statistics generated by X_1, X_2, \dots, X_n , which is defined with probability 1 (c.f., Bairamov and Petunin (1991)), and $P\{X_{n+1} \in (X_{(i)}, X_{(j)})\} = (j - i)/(n + 1)$.

Let $\{X_n\}_{n \geq 1}$ be a sequence of i.i.d. random variables with continuous d.f. $F(x)$. Let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ be the almost sure (a.s.) defined order statistics generated by X_1, X_2, \dots, X_n . Consider the spacings $X_{(1)} - X_{(0)}, X_{(2)} - X_{(1)}, X_{(3)} - X_{(2)}, \dots, X_{(n)} - X_{(n-1)}$ with $X_{(0)} = \alpha(F) \equiv \inf\{x : F(x) > 0\}$. Define a random variable ν as follows: $\nu = k$ iff $X_{(k)} - X_{(k-1)} \leq X_{(i)} - X_{(i-1)}, i = 1, 2, \dots, n$. It is clear that ν is the index of a spacing having minimal length. Bairamov and Eryilmaz (2000) show that if $F(x) = 1 - \exp(-\lambda x), x \geq 0, \lambda > 0$, then, for $s = 0, 1, \dots, m$,

$$\begin{aligned}
& P\{X_{n+1}, X_{n+2}, \dots, X_{n+s} \in (X_{(\nu-1)}, X_{(\nu)}), X_{n+s+1}, \dots, X_{n+m} \notin (X_{(\nu-1)}, X_{(\nu)})\} \\
&= \frac{2}{n+1} \sum_{i=0}^{m-s} (-1)^i \binom{m-s}{i} \frac{n+s+i+1}{s+i+2} \left(\binom{\frac{n(n+1)}{2} + s + i}{s+i} \right)^{-1} \quad (1) \\
& s = 0, 1, 2, \dots, m,
\end{aligned}$$

where $X_{(0)} = 0$. Let S_m be the number of $k, k \in \{n+1, \dots, n+m\}$, such that $X_k \in (X_{(\nu-1)}, X_{(\nu)})$. Denote by $A(m, n, s)$ the right hand side (r.h.s.)

in (1). An almost straightforward consequence of (1) is that

$$P\{S_m = s\} = \binom{m}{s} A(m, n, s), \quad s = 0, 1, 2, \dots, m. \quad (2)$$

2. ASYMPTOTIC DISTRIBUTION OF STATISTICS BASED ON MINIMAL SPACING

Let X_1, X_2, \dots, X_n be an independent copies of nonnegative random variable having continuous distribution with d.f. F . Also, let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ be order statistics constructed on the basis of X_1, X_2, \dots, X_n . Consider the spacings $X_{(1)} - X_{(0)}, X_{(2)} - X_{(1)}, X_{(3)} - X_{(2)}, \dots, X_{(n)} - X_{(n-1)}$ with $X_{(0)} = 0$. Define a random variable ν as follows: $\nu = k$ iff $X_{(k)} - X_{(k-1)} \leq X_{(i)} - X_{(i-1)}, i = 1, 2, \dots, n$. It is clear that ν is the index of a spacing having minimal length. Let Y_1, Y_2, \dots, Y_m be another sample from nonnegative random variable Y with continuous d.f. G .

It is clear that $S_m(r, s)$ denotes the number of X'_i 's falling above the threshold $X_{u(r)}$ and below the threshold $X_{u(s)}$.

Let S_m be a number of observations Y_1, Y_2, \dots, Y_m falling above the random threshold $X_{(\nu-1)}$ and below the threshold $X_{(\nu)}$. It is clear that in case $F = G$ the second sample Y_1, Y_2, \dots, Y_m can be interpreted as a continuation $X_{n+1}, X_{n+2}, \dots, X_{n+m}$ of the first sample X_1, X_2, \dots, X_n .

Theorem 1. The asymptotic distribution of $\frac{S_m}{m}$ for large m is

$$\lim_{m \rightarrow \infty} \sup_{0 \leq x \leq 1} \left| P \left\{ \frac{S_m}{m} \leq x \right\} - P \left\{ G(X_{(\nu)}) - G(X_{(\nu-1)}) \leq x \right\} \right| = 0.$$

Proof. By definition

$$S_m = \sum_{i=1}^m \xi_i = \sum_{i=1}^m I_{(X_{(\nu-1)}, X_{(\nu)})}(X_{n+i}), \quad (3)$$

where $I_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$. Using the representation (3) one has

$$\begin{aligned} P \left\{ \frac{S_m}{m} \leq x \right\} &= P \left\{ \frac{1}{m} \sum_{i=1}^m I_{(X_{(\nu-1)}, X_{(\nu)})}(X_{n+i}) \leq x \right\} \\ &= P \left\{ \int_{-\infty}^{\infty} I_{(X_{(\nu-1)}, X_{(\nu)})}(u) dG_m^*(u) \leq x \right\} \end{aligned} \quad (4)$$

where $G_m^*(u)$ is the empirical distribution function constructed by the sample $X_{n+1}, X_{n+2}, \dots, X_{n+m}$ of size m . Note that G_m^* and X_1, X_2, \dots, X_n are independent random variables and $X_1(\omega), X_2(\omega), \dots, X_n(\omega), X_{n+1}(\omega), X_{n+2}(\omega), \dots, X_{n+m}(\omega), \dots$ is considered as a sequence of i.i.d. random variables defined in probability space $\{\Omega, \mathfrak{F}, P\}$, where Ω is a set of points, \mathfrak{F} is a σ -field of subsets of Ω , and P is a probability measure given in \mathfrak{F} . Denote

$$H^*(G) = \int_{-\infty}^{\infty} I_{(x,y)}(u) dG(u), \quad (x < y, x, y \text{ are fixed})$$

$$H(G) = \int_{-\infty}^{\infty} I_{(X_{(k-1)}, X_{(k)})}(u) dG(u).$$

$H(G) = H(G)(\omega)$ is a random variable defined in a probability space $\{\Omega, \mathfrak{F}, P\}$. By using Glivenko-Cantelli theorem one can observe that $H^*(G_m^*) \rightarrow H^*(G)$ almost surely. From (4) it follows that

$$\begin{aligned} P \left\{ \frac{S_m}{m} \leq x \right\} &= \sum_{k=1}^n P \left\{ \int_{-\infty}^{\infty} I_{(X_{(k-1)}, X_{(k)})}(u) dG_m^*(u) \leq x, \nu = k \right\} \\ &= \sum_{k=1}^n P \left\{ \int_{-\infty}^{\infty} I_{(X_{(k-1)}, X_{(k)})}(u) dG_m^*(u) \leq x, X_{(k)} - X_{(k-1)} \leq X_{(i)} - X_{(i-1)}, \right. \\ i = 1, 2, \dots, n \} &= \sum_{k=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^y P \left\{ \int_{-\infty}^{\infty} I_{(X_{(k-1)}, X_{(k)})}(u) dG_m^*(u) \leq x, X_{(k)} - X_{(k-1)} \leq \right. \\ &\leq X_{(i)} - X_{(i-1)}, i = 1, 2, \dots, n \mid X_{(k-1)} = x, X_{(k)} = y \} dF_{k-1,k}(x, y), \end{aligned} \quad (5)$$

where $F_{k-1,k}(x, y)$ is the joint distribution function of $(X_{(k-1)}, X_{(k)})$. After simplifying (5) it follows that

$$P \left\{ \frac{S_m}{m} \leq x \right\} =$$

$$\begin{aligned}
& \sum_{k=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^y P \left\{ \int_{-\infty}^{\infty} I_{(x,y)}(u) dG_m^*(u) \leq x, x-y \leq X_{(i)} - X_{(i-1)}, i = 1, 2, \dots, n, \right. \\
& \quad \left. X_{(k-1)} = x, X_{(k)} = y \right\} \frac{1}{P \{X_{(k-1)} = x, X_{(k)} = y\}} dF_{k-1,k}(x, y) \\
&= \sum_{k=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^y P \left\{ \int_{-\infty}^{\infty} I_{(x,y)}(u) dG_m^*(u) \leq x \right\} P \{x-y \leq X_{(i)} - X_{(i-1)}, i = 1, 2, \dots, n, \\
& \quad X_{(k-1)} = x, X_{(k)} = y\} \frac{1}{P \{X_{(k-1)} = x, X_{(k)} = y\}} dF_{k-1,k}(x, y) \stackrel{a.s.}{\rightarrow} \\
& \stackrel{a.s.}{\rightarrow} \sum_{k=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^y P \left\{ \int_{-\infty}^{\infty} I_{(x,y)}(u) dG(u) \leq x \right\} P \{x-y \leq X_{(i)} - X_{(i-1)}, i = 1, 2, \dots, n, \\
& \quad X_{(k-1)} = x, X_{(k)} = y\} \frac{1}{P \{X_{(k-1)} = x, X_{(k)} = y\}} dF_{k-1,k}(x, y) \\
&= \sum_{k=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^y P \left\{ \int_{-\infty}^{\infty} I_{(x,y)}(u) dG(u) \leq x, x-y \leq X_{(i)} - X_{(i-1)}, i = 1, 2, \dots, n, \right. \\
& \quad \left. X_{(k-1)} = x, X_{(k)} = y \right\} \frac{1}{P \{X_{(k-1)} = x, X_{(k)} = y\}} dF_{k-1,k}(x, y) \\
&= \sum_{k=1}^n P \left\{ \int_{-\infty}^{\infty} I_{(X_{(k-1)}, X_{(k)})}(u) dG(u) \leq x, X_{(k)} - X_{(k-1)} \leq X_{(i)} - X_{(i-1)}, i = 1, 2, \dots, n \right\} \\
&= P \left\{ \int_{-\infty}^{\infty} I_{(X_{(\nu-1)}, X_{(\nu)})}(u) dG(u) \leq x \right\} = P \{G(X_{(\nu)}) - G(X_{(\nu-1)}) \leq x\}.
\end{aligned}$$

(Note that expressions of type $\frac{1}{P \{X_{(k-1)} = x, X_{(k)} = y\}}$ above should be understood as

$$\lim_{h \rightarrow 0} \frac{1}{P \{x \leq X_{(k-1)} < x+h, y \leq X_{(k)} < y+h\}}.$$

Theorem thus proved.

Remark 1. Let $F(x) = G(x) = 1 - \exp(-\lambda x)$, $x \geq 0$. Then

$$P \{F(X_{(\nu)}) - F(X_{(\nu-1)}) \leq x \mid \nu = 1\} = 1 - (1-x)^{\frac{n(n+1)}{2}}, 0 \leq x \leq 1.$$

Proof. In fact by definition of ν one has

$$P \{F(X_{(\nu)}) - F(X_{(\nu-1)}) \leq x \mid \nu = 1\} = \frac{P \{F(X_{(\nu)}) - F(X_{(\nu-1)}) \leq x, \nu = 1\}}{P \{\nu = 1\}}.$$

Consider

$$\begin{aligned} P \{F(X_{(\nu)}) - F(X_{(\nu-1)}) \leq x, \nu = 1\} &= P \{F(X_{(1)}) - F(X_{(0)}) \leq x, \nu = 1\} \\ &= P \{F(X_{(1)}) \leq x, X_{(1)} \leq X_{(2)} - X_{(1)}, \dots, X_{(1)} \leq X_{(n)} - X_{(n-1)}\} \\ &= P \{F(X_{(1)}) \leq x, X_{(1)} \leq X_{(2)} - X_{(1)}, \dots, X_{(1)} \leq X_{(n)} - X_{(n-1)} \mid X_{(1)} = t\} dF_1(t) \\ &= n\lambda_{F(t) \leq x} (e^{-\lambda t})^{\frac{n(n+1)}{2}} dt = n\lambda_0^{-\frac{1}{\lambda} \ln(1-x)} (e^{-\lambda t})^{\frac{n(n+1)}{2}} dt \\ &= \frac{2}{n+1} (1 - (1-x)^{\frac{n(n+1)}{2}}) \quad , \quad 0 \leq x \leq 1. \end{aligned}$$

After some algebra, using $P \{\nu = 1\} = \frac{2}{n+1}$ one obtains

$$P \{F(X_{(\nu)}) - F(X_{(\nu-1)}) \leq x \mid \nu = 1\} = 1 - (1-x)^{\frac{n(n+1)}{2}} \quad , \quad 0 \leq x \leq 1.$$

Lemma 1. Let X_1, X_2, X_3 i.i.d. random variables with continuous d.f. F . Let $X_{(\nu)} - X_{(\nu-1)} \leq X_{(i)} - X_{(i-1)}$, $i = 2, 3$. The p.d.f. of random variable $F(X_{(\nu)}) - F(X_{(\nu-1)})$ is

$$\begin{aligned} f^*(t) &= 6 \int_t^1 [1 - F(2F^{-1}(v) - F^{-1}(v-t)) + F(2F^{-1}(v-t) - F^{-1}(v))] dv \\ & \quad 0 < t < 1. \end{aligned}$$

Proof. The d.f. of random variable $F(X_{(\nu)}) - F(X_{(\nu-1)})$ can be found as follows :

$$F^*(t) = P \{F(X_{(\nu)}) - F(X_{(\nu-1)}) \leq t\} = \iint_C f_{\nu-1, \nu}(x, y) dx dy \quad (6)$$

where $C = \{(x, y) : x < y, F(y) - F(x) \leq t\}$ and $f_{\nu-1, \nu}(x, y)$ is the joint p.d.f. of $X_{(\nu-1)}$ and $X_{(\nu)}$. It is known that

$$f_{\nu-1, \nu}(x, y) = 6f(x)f(y) [1 - F(2y-x) + F(2x-y)] \quad , \quad x < y \quad (7)$$

(c.f., David (1981)).

Let us take in (6) $F(x) = u$ and $F(y) = v$. The Jacobian of this transformation is equal to $(f(F^{-1}(u))f(F^{-1}(v)))^{-1}$. Then from (6) using (7) one has

$$F^*(t) = 6 \int_0^t \int_0^v [1 - F(2F^{-1}(v) - F^{-1}(u)) + F(2F^{-1}(u) - F^{-1}(v))] dudv$$

$$+ 6 \int_t^1 \int_{v-t}^v [1 - F(2F^{-1}(v) - F^{-1}(u)) + F(2F^{-1}(u) - F^{-1}(v))] dudv.$$

The proof is completed by differentiating $F^*(t)$ with respect to t .

Example. Let X_1, X_2, X_3 be a sample from uniform on $(0, 1)$ distribution. Then

$$f^*(t) = \begin{cases} 6(2t - 1)^2, & 0 < t < \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}, \quad (8)$$

In fact, from Lemma 1 we have

$$f^*(t) = 6 \int_t^1 dv - 6 \int_t^1 F(v+t)dv + 6 \int_t^1 F(v-2t)dv. \quad (9)$$

It is evident that

$$F(v+t) = \begin{cases} 0 & \text{if } v \leq -t \\ v+t & \text{if } -t < v < 1-t \\ 1 & \text{if } v \geq 1-t \end{cases}, F(v-2t) = \begin{cases} 0 & \text{if } v \leq 2t \\ v-2t & \text{if } 2t < v < 1+2t \\ 1 & \text{if } v \geq 1+2t \end{cases}.$$

Let $0 < t < \frac{1}{2}$. Then $t < 1-t$ and one has

$$6 \int_t^1 F(v+t)dv = 6 \int_t^{1-t} (v+t)dv + 6 \int_{1-t}^1 dv = 3 - 12t^2 + 6t, \quad (10)$$

$$6 \int_t^1 F(v-2t)dv = 6 \int_{2t}^1 (v-2t)dv = 3 - 12t + 12t^2.$$

Let $\frac{1}{2} < t < 1$. Then $t > 1 - t$, $F(v + t) = 1$, $F(v - 2t) = 0$ and one has

$$\begin{aligned} 6 \int_t^1 F(v + t) dv &= 6 \int_t^1 dv = 6(1 - t), \\ 6 \int_t^1 F(v - 2t) dv &= 0. \end{aligned} \quad (11)$$

Using (10) and (11) in (9) one obtains (8).

It is clear that in this case the d.f. of $F(X_{(\nu)}) - F(X_{(\nu-1)})$ is

$$F^*(t) = \begin{cases} 8t^3 - 12t^2 + 6t, & 0 < t < 1/2 \\ 1, & t \geq 1/2 \end{cases}$$

Lemma 2. Let X_1, X_2, \dots, X_n i.i.d r.v.'s with d.f. $F(x) = 1 - \exp\{-\lambda x\}$, $x > 0$, $\lambda > 0$ and $(X_{(\nu-1)}, X_{(\nu)})$ is minimal spacing of this sample. In this case the d.f. of random variable $X_{(\nu)} - X_{(\nu-1)}$ is

$$P\{X_{(\nu)} - X_{(\nu-1)} \leq x\} = 1 - \exp\left\{-\frac{n(n+1)}{2}\lambda x\right\}, \quad x > 0.$$

Remark 2. Let X_1, X_2, \dots, X_n i.i.d r.v.'s with d.f. F . If F is exponential d.f. with parameter λ then

$$\frac{n+1}{2} (X_{(\nu)} - X_{(\nu-1)}) \stackrel{d}{=} X_{(1)}$$

3. RECORD STATISTICS

Let $\{X_n\}_{n \geq 1}$ be a sequence of i.i.d r.v.'s with continuous distribution function F . Define a record times of this sequence as follows: $u(1) = 1$, $u(n) = \min j : j > u(n-1), X_j > X_{u(n-1)}, n > 1$. Let $X_{u(1)}, X_{u(2)}, \dots$ be corresponding record values.

The d.f. and probability density function (p.d.f.) of record values can be expressed in terms of

$$R(x) = -\ln(1 - F(x)) \quad \text{and} \quad r(x) = \frac{d}{dx} R(x) = \frac{f(x)}{1 - F(x)}.$$

It is known that the distribution of n th record value is

$$F_n(x) = P \{X_{u(n)} \leq x\} = \int_{-\infty}^x \frac{R^{n-1}(u)}{(n-1)!} dF(u) \quad , -\infty < x < \infty.$$

The joint p.d.f. of $X_{u(i)}$ and $X_{u(j)}$ is

$$f(x_i, x_j) = \frac{(R(x_i))^{i-1}}{(i-1)!} r(x_i) \frac{(R(x_j) - R(x_i))^{j-i-1}}{(j-i-1)!} f(x_j) \quad , -\infty < x_i < x_j < \infty$$

The details of record theory can be found in Galambos (1987), Nagaraja (1988), Nevzorov (1988), Ahsanullah (1995), Arnold, Balakrishnan and Nagaraja (1998) among others.

Let $X_{u(n)}$ be the n th record value of a sequence of i.i.d. r.v.'s.

Suppose that $X_{u(n)+1}, X_{u(n)+2}, \dots, X_{u(n)+m}$ be the m observations that succeed $X_{u(n)}$.

For $i = 1, 2, \dots, m$ and $1 \leq r < s \leq n$, it is known that

$$P \{X_{u(n)+i} \in (X_{u(r)}, X_{u(s)})\} = \frac{1}{2^r} - \frac{1}{2^s}$$

(c.f., Bairamov (1997)).

Define the random variables $\xi_1, \xi_2, \dots, \xi_m$ as follows:

$$\xi_i = \begin{cases} 1 & \text{if } X_{u(n)+i} \in (X_{u(r)}, X_{u(s)}) \\ 0 & \text{if } X_{u(n)+i} \notin (X_{u(r)}, X_{u(s)}) \end{cases} \quad , \quad i = 1, 2, \dots, m \quad , \quad r < s.$$

Denote

$$\nu_m = \sum_{i=1}^m \xi_i$$

ν_m is the number of observations $X_{u(n)+1}, X_{u(n)+2}, \dots, X_{u(n)+m}$ falling into interval $(X_{u(r)}, X_{u(s)})$. It is clear that the r.v.'s $\xi_1, \xi_2, \dots, \xi_m$ are dependent. Bairamov and Eryilmaz (2000) show that for $r = k - 1$ and $s = k$ ($k = 2, \dots, n$)

$$P \{\nu_m = j\} = \binom{m}{j} \sum_{i=0}^{m-j} (-1)^i \binom{m-j}{i} \frac{1}{(j+i+1)^k}$$

$$j = 0, 1, \dots, m.$$

Let $\{X_n\}_{n \geq 1}$ be a sequence of i.i.d. r.v.'s with continuous d.f. F , $X_{u(1)}, X_{u(2)}, \dots$ is a sequence of record values. Let X'_1, X'_2, \dots, X'_m be i.i.d. random variables with continuous d.f. G . Define the following random variables

$$\xi_i^*(r, s) = \begin{cases} 1 & \text{if } X'_i \in (X_{u(r)}, X_{u(s)}) \\ 0 & \text{if } X'_i \notin (X_{u(r)}, X_{u(s)}) \end{cases}, \quad i = 1, 2, \dots, m, \quad r < s.$$

and let

$$S_m(r, s) = \sum_{i=1}^m \xi_i^*(r, s).$$

It is clear that $S_m(r, s)$ denotes the number of X'_i 's falling above the threshold $X_{u(r)}$ and below the threshold $X_{u(s)}$. It is true that

$$\lim_{m \rightarrow \infty} \sup_{0 \leq x \leq 1} \left| P \left\{ \frac{S_m(r, s)}{m} \leq x \right\} - P \{ G(X_{u(s)}) - G(X_{u(r)}) \leq x \} \right| = 0.$$

The case $F = G$ is of interest on its own and we have the following results for this case:

Let $F = G$. Then

$$\lim_{m \rightarrow \infty} \sup_{0 \leq x \leq 1} \left| P \left\{ \frac{S_m(r, s)}{m} \leq x \right\} - P \{ F(X_{u(s)}) - F(X_{u(r)}) \leq x \} \right| = 0.$$

where $U_{rs} = F(X_{u(s)}) - F(X_{u(r)})$ has d.f.

$$\begin{aligned} P \{ U_{rs} \leq x \} &= \frac{1}{(r-1)!(s-r-1)!} \int_0^x \int_0^{1-t_1} \left[\ln \frac{1}{1-t_2} \right]^{r-1} \frac{1}{1-t_2} \times \\ &\quad \times \left[\ln \frac{1-t_1}{1-t_1-t_2} \right]^{s-r-1} dt_2 dt_1, \end{aligned}$$

and p.d.f.

$$\begin{aligned} f_{U_{rs}}(x) &= \frac{1}{(r-1)!(s-r-1)!} \int_0^{1-x} \left[\ln \frac{1}{1-t_2} \right]^{r-1} \frac{1}{1-t_2} \left[\ln \frac{1-x}{1-x-t_2} \right]^{s-r-1} dt_2, \\ 0 &< x < 1. \end{aligned}$$

Now let again $F = G$ and $r = k-1, s = k$ ($k = 2, \dots, n$). Then the expected value and the variance of $S_m(r, s)$ is:

$$ES_m(k-1, k) = m/2^k,$$

$$VarS_m(k-1, k) = m(1/2^k - 1/3^k) + m^2(1/3^k - 1/2^{2k}).$$

Let $F = G$. For $r = k-1$ and $s = k$ ($k = 2, 3, \dots, n$). Then

$$\lim_{m \rightarrow \infty} \sup_{0 \leq x \leq 1} \left| P \left\{ \frac{S_m(r, s)}{m} \leq x \right\} - P \{U_{k-1, k} \leq x\} \right| = 0,$$

where $U_{k-1, k} = F(X_{u(k)}) - F(X_{u(k-1)})$ has d.f.

$$D_k(x) \equiv P \{U_{k-1, k} \leq x\} = \frac{1}{(k-1)!} \int_0^x \left[\ln \frac{1}{u} \right]^{k-1} du, \quad 0 < x < 1.$$

and p.d.f.

$$d_k(x) \equiv f_{U_{k-1, k}}(x) = \frac{1}{(k-1)!} \left[\ln \frac{1}{x} \right]^{k-1}, \quad 0 < x < 1.$$

Also it is true that

$$P \left\{ \frac{S_m(k-1, k) - ES_m(k-1, k)}{\sqrt{VarS_m(k-1, k)}} \leq x \right\} \xrightarrow{m \rightarrow \infty} D_k(ax + b),$$

where

$$a = \sqrt{\frac{1}{3^k} - \frac{1}{2^{2k}}} \quad \text{and} \quad b = \frac{1}{2^k}.$$

Remark 3. Let $\{X_n\}_{n \geq 1}$ be a sequence of i.i.d. r.v.'s with continuous d.f. F , $X_{u(1)}, X_{u(2)}, \dots$ be corresponding upper record values and $X_{L(1)}, X_{L(2)}, \dots$ be corresponding lower record values of this sequence. If F is uniform d.f. on $(0, 1)$ then

$$X_{L(k)} \stackrel{d}{=} X_{u(k)} - X_{u(k-1)}$$

References

Ahsanullah, M.(1995) *Record Statistics*. Nova Science publishers, Inc., Commack, NY.

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

Bairamov, I.G. and Petunin, Yu.I. (1991) Structure of invariant confidence intervals containing the main distributed mass. *Theor. Prob. Appl.* Vol **35**, N_1, P.15-26.

Bairamov, I.G. (1997) Some distribution free properties of statistics based on record values and characterizations of the distributions through a record. *Journal of Applied Statistical Science.* Vol. **5**, N_1, P.17-25.

Bairamov, I.G. and Gebizlioglu, O.L. (1998) Asymptotic distributions of the statistics based on order statistics, record values and invariant confidence intervals. *Proceedings of International conference "Asymptotic Methods in Probability and Mathematical Statistics"*, St: Petersburg, June 24-28, 1998.

Bairamov, I.G., Gebizlioglu, O.L. and Kaya, M.

Bairamov, I.G. and Eryilmaz S.N. (2000) Distributional Properties of Statistics Based on Minimal Spacing and Record Exceedance Statistics. *Journal of Statistical Planning and Inference.* (in press)

David, H.A.(1981) *Order Statistics.* Wiley. New York.

Epstein, B. (1954) Tables for the distribution of the number of exceedances. *Ann. Math.. Statist.* **25**, 762-768.

Galambos, J. (1987) *The Asymptotic Theory of Extreme Order Statistics.*(2nd ed). Klieger, Malabar (FL).

Johnson, N.L., Kotz, S. and Kemp, A. (1992) *Univariate discrete distributions.* Wiley, New York.

Leadbetter, M.L., Lindgren, G. and Rootzen, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New-York.

Nagaraja, H.N.(1988) Record values and related statistics. A Review. *Commun in Statist. Theor Meth.* **17**, 2223-2238.

Nevzorov, V.B.(1988) Records. *Theor.Prob.Apl.* **32**, 201-228.

Robbins, H. (1944) On distribution-free tolerance limits in random sampling. *Ann. Math. Statist.*, **15**, 214-216.

Siddiqui, M.M. (1970) Order statistics for a sample and extended sample. *Nonparametric Techniques in Statistical Inference.* (Ed. M.L.Puri), 417-423, Cambridge University Press, Cambridge.

Shewart W.A. (1931) *Economic control of quantity of manufactured product.* D. Van Nostrand, Princeton, N.J.

Wilks, S.S. (1941) On the determination of sample sizes for setting tolerance limits. *Ann. Math.Stat.* Vol. **12**, 91-96.

Wilks, S.S. (1942) Statistical prediction with special references to the problem of tolerance limits. *Ann. Math.Stat.* Vol. **13**, 400-409.

Liter. extended

Bairamov, I.G., Petunin, Yu.I. (1990) Structure of invariant confidence intervals containing the main distributed mass. *Theorya Veroyatnostey i ee Primeneniya*. Vol **35**, No 1, P.15-26. (English translation in *Theor. Probab. Appl.* 1991, Vol 35, No 1, P.15-26.)

David, H. A. (1981) *Order Statistics*. Second edition, John Wiley and Sons, New York.

Gumbel, E.J. and Schelling, H, von (1950) The distribution of the number of exceedances. *Ann. Math.. Statist.* **21**, 247-262.

Johnson, N. and Kotz, S. (1991) Some generalizations of Bernoulli and Polya-Eggenberger contagion models. *Statist. Papers* **32**, P.1-17.

Johnson, N. and Kotz, S. (1994) Further comments on Matveychuk and Petunin's generalized Bernoulli model, and nonparametric tests of homogeneity. *J. Statist. Plann. Infer.* **41** , P. 61-72.

Katzenbeisser, W. (1985) The distribution of two sample location exceedance test statistics under Lehmann alternatives. *Statist. Hefte (now Papers)* **26**,131-138.

Katzenbeisser, W. (1986) The exact power of two-sample location tests based on exceedances against shift alternatives. *Math. Op. Forsch. (Statistics)* **20**, 47-54.

Matveychuk, S.A., Petunin Yu.I.(1990/1991) A generalization of the Bernoulli model arising in order statistics. I/II *Ukr. Math. J.* Vol. **42**, P. 518-528/ **43**, P. 779-785.

Rudin W. (1959) *Principles of Mathematical Analysis*. Mc Graw-Hill Book Company, INC. New York.

Sarkadi, K. (1957) On the distribution of the number of exceedances. *Ann. Math. Stat.*, **28**, 1021-1022.

Todhunter, I. (1865) *A History of the Mathematical Theory of Probability*. Cambridge: Macmillan (Reprinted 1965, New York: Chelsea).

Wesolowsky, J. and Ahsanullah, M. (1998) Distributional properties of exceedance statistics. *Ann. Inst. Statist. Math.* Vol. **50**, N_3, 543-565.

Bairamov I. and Kotz S. (2000) On distributions of exceedances associated with order statistics and record values for arbitrary distributions. *Statistical Papers*. (in press)