Advances in exceedance statistics based on ordered random variables

Ismihan Bairamov
Department of Mathematics, Izmir University of Economics
Sakarya Caddesi 156, 35330 Balcova, Izmir
e-mail: ismihan.bayramoglu@ieu.edu.tr

Summary. This review paper consists of a description of the recent results connected with the exceedance statistics based on ordered random variables. In particular, the distributions of exceedance statistics in general random threshold model based on functions of the independent and identically distributed random variables are given.

Introduction

Let X and Y be two random variables with continuous distribution functions F and Q and $X_1, X_2, ..., X_n, ...$ and $Y_1, Y_2, ..., Y_m, ...$ be independent copies of X and Y respectively. Consider two Borel functions $f_1(u_1, u_2, ..., u_n)$ and $f_2(u_1, u_2, ..., u_n)$ with the following property:

$$f_1(u_1, u_2, ..., u_n) \leq f_2(u_1, u_2, ..., u_n) \qquad (u_1, u_2, ..., u_n) \in R^n. \tag{1}$$

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

$$\xi_i = \begin{cases} 1, & if \quad Y_i \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n)) \\ 0, & \text{otherwise} \end{cases} .$$

$i = 1, 2.., n$. We call $f_1(X_1, X_2, ..., X_n)$ and $f_2(X_1, X_2, ..., X_n)$ the lower and upper random thresholds, respectively. Consider

$$\nu_m = \sum_{i=1}^{n} \xi_i.$$

It is clear that the random variable ν_m shows the number of the elements of the sample $Y_1, Y_2, ..., Y_m$ falling into random interval $(f_1(X_1, X_2, ..., X_n)$, $f_2(X_1, X_2, ..., X_n)$. We call ν_m the exceedance statistics in independent and identically distributed (iid) sequences of observations. We are interested in distributional properties of ν_m. The particular cases for the lower and upper thresholds being the r^{th} and s^{th} order statistics, $X_{r:n}$ and $X_{s:n}$ from the sample $X_1, X_2, ..., X_n$ were investigated in connection with the theory of tolerance limits and invariant confidence intervals for future observations. See for instance Wilks(1941) , Robbins (1944), Gumbel and Shelling (1950), Epstein (1954), Sarkadi (1957), Siddigui (1970), David (1981), Bairamov and Petunin (1991).

Some of the results on this topic were used to construct a statistical criteria for testing hypothesis $H_0 : F = Q$ against some classes of alternatives. For example Matveychuk and Petunin (1991) and Johnson and Kotz (1991) studied a generalized Bernoulli model defined in terms of placement statistics from two random samples. Katzenbeisser (1985), obtained a formula for the distribution of ν_m when $f_1(X_1, X_2, ..., X_n) = -\infty$ and $f_2(X_1, X_2, ..., X_n) = X_{r:n}$ and proposed a test criterion for testing the null hypothesis $H_0 : F(x) = Q(x)$ versus Lehmann alternatives $Q(x) = [F(x)]^{\theta}, \theta \neq 1$. He extended these results to shift alternatives (Katzenbeisser (1986)). Matveychuk and Petunin (1991), Johnson and Kotz (1991), (1994) investigated the test criterion for testing the hypothesis $H_0 : F(x) = Q(x)$ by using ν_m.

1. Exceedance statistics in iid sequences of observations from continuous distributions

The probability $P\{Y \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\}$ plays an important role in determinig of distributions of exceedance statistics. Especially, it is desirible that this probability would not depend on the distribution function if the hypothesis $H_0 : F = Q$ is true. For the general consideretation assume that the distribution function F_X belongs to some general class of distributions $\Im$. Let X_{n+1} be the next $(n+1)^{th}$ observation which is independent of $X_1, X_2, ..., X_n$. We say that the random interval $(f_1(X_1, X_2, ..., X_n),\ f_2(X_1, X_2, ..., X_n))$ containing the future observation is invariant (or distribution free) with respect to the class $\Im$ if the probability $p = P\{X_{n+1} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\}$ is the same for all distributions from the class $\Im$. It is easy to show that the order statistics $X_{r:n}$ and $X_{s:n}$, $1 \leq r < s \leq n$ form an invariant confidence interval containing the future observation for the class of all continuous distribution functions and in this case $p = \frac{s-r}{n+1}$. It is also known that if f_1 and f_2 are continuous and symmetric functions of n variables then the order statistics are unique invariant intervals for the class $\Im_c$ (Bairamov and Petunin (1991)). It is interesting to know that if one narrows the class $\Im_c$ to some parametric subclass $\wp = \{P_\theta, \theta \in \Theta\} \subset \Im_c$, then there exist the invariant intervals for the class $\wp$ which are different those constructed by the order statistics. More precisely, let $X_{n+1}, X_{n+2}, ..., X_{n+m}$ be a new sample independent of $X_1, X_2, ..., X_n$. Then we can write for $\theta \in \Theta$

$$P_\theta\{X_{n+1}, ..., X_{n+m} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\} =$$

$$= \int ... \int [F_\theta(f_2(u_1, ..., u_n)) - F_\theta(f_1(u_1, ..., u_n))]^m \, dF_\theta(u_1)...dF_\theta(u_n) =$$

$$= E_\theta [F_\theta(f_2(X_1, X_2, ..., X_n)) - F_\theta(f_1(X_1, X_2, ..., X_n))]^m .$$

Denote

$$T_n(X_1, X_2, ..., X_n, \theta) = F_\theta(f_2(X_1, X_2, ..., X_n)) - F_\theta(f_1(X_1, X_2, ..., X_n)) \text{ and}$$
$$G_\theta(u) = P_\theta \left\{T_n(X_1, X_2, ..., X_n, \theta) \le u\right\}.$$

We can formulate the following

Theorem 1.1. Bairamov et.al (1999) If the distribution of the random variable (r.v.) $T_n(X_1, X_2, ..., X_n, \theta)$ is the same for all $\theta \in \Theta$ (i.e. the d.f of S_n is independent from θ), then $(f_1(X_1, X_2, ..., X_n),\ f_2(X_1, X_2, ..., X_n))$ is an invariant confidence interval for family $\wp$.

Theorem 1.1 pave a way of methods for constructing invariant confidence intervals. Development of these methods can be extended to families of distribution with location parameters. Let us assume that we have the family of distributions $\wp = \{F_\theta(x) = F(x-\theta), \theta \in \Theta\}$, F is known. If it is true that $X_1, X_2, ..., X_n$ has d.f. $F_\theta \in \wp$, then one can write

$$\overline{T_n}(X_1, X_2, ..., X_n, \theta) = F_\theta(f_2(X_1, X_2, ..., X_n)) - F_\theta(f_1(X_1, X_2, ..., X_n))$$

$$= F(f_2(X_1, X_2, ..., X_n) - \theta) - F(f_1(X_1, X_2, ..., X_n) - \theta)$$

Define $D^+ = \{(u_1, u_2, ..., u_n)\ ;\ u_1 \ge u_2 \ge ... \ge u_n\}$and let $a = (a_1, a_2, ..., a_n) \in R^n$, $b = (b_1, b_2, ..., b_n) \in R^n$, $a_{[1]} \ge a_{[2]} \ge ... \ge a_{[n]}$, $b_{[1]} \ge b_{[2]} \ge ... \ge b_{[n]}$ where $a_{[.]}$, $b_{[.]}$ designate order of magnitude. Given these definitions, under the conditions of

$$1. \sum_{i=1}^{n} a_{[i]} = \sum_{i=1}^{n} b_{[i]}$$

$$2. \sum_{i=1}^{k} a_{[i]} \le \sum_{i=1}^{k} b_{[i]}, k = 1, 2, ..., n-1$$

it is said that vector a is majorant to vector b. This is expressed symbolically as $a \prec b$ (see Marshal and Olkin, 1979).

It is known that the necessary and sufficient condition for $a \prec b$ to hold true is the condition that

$$\sum_{i=1}^{n} a_i u_i \le \sum_{i=1}^{n} b_i u_i, \text{for all } u = (u_1, u_2, ..., u_n) \in D^+.$$

(see Marshal and Olkin,1979, Chapter 4). In order to utilize this theorem let $a = (\frac{1}{n}, \frac{1}{n}, ... \frac{1}{n})$, $b = (\frac{1}{n-1}, \frac{1}{n-1}, ... \frac{1}{n-1}, 0)$, and let $X_{[1]}, X_{[2]}, ..., X_{[n]}$ be the order statistics for the sample $X_1, X_2, ..., X_n$, $X_{[n-i+1]} = X_{(i)}$. Set

$$f_1(X_1, X_2, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_{[i]}\ ,\ f_2(X_1, X_2, ..., X_n) = \frac{1}{n-1}\sum_{i=1}^{n-1} X_{[i]}.$$

Hence , it follows that

$$F_\theta\left(\frac{1}{n-1}\sum_{i=1}^{n-1}X_{[i]}\right)-F_\theta\left(\frac{1}{n}\sum_{i=1}^{n}X_{[i]}\right)$$

$$=F\left(\frac{1}{n-1}\sum_{i=1}^{n-1}(X_{[i]}-\theta)\right)-F\left(\frac{1}{n}\sum_{i=1}^{n}(X_{[i]}-\theta)\right)$$

$$=F\left(\frac{1}{n-1}\sum_{i=1}^{n-1}(X_{(n-i+1)}-\theta)\right)-F\left(\frac{1}{n}\sum_{i=1}^{n}(X_{(n-i+1)}-\theta)\right).$$

It is obvious that, here the distribution of $X_{(n-i+1)}-\theta$ is independent of θ ; which means that distribution of r.v. $\overline{T_n}$ is the same for all elements of class $\wp$. Obviously, it will be true that for a two parameter family of distributions

$$\wp_1=\left\{F_{\theta,\mu}(x)=F(\frac{x-\mu}{\theta})\ ,\ \theta\in\Theta\ ,\ \mu\in\Theta_1\ ,\ F(x)\text{ is known}\right\}$$

the distribution of a similar random variable

$$S_n^*\left(X_1,X_2,...,X_n,\theta,\mu\right)=F_{\theta,\mu}\left(\frac{1}{n-1}\sum_{i=1}^{n-1}(X_{[i]})\right)-F_{\theta,\mu}\left(\frac{1}{n}\sum_{i=1}^{n}(X_{[i]})\right),$$

is also independent from θ and μ.

The exponential distribution case

The exponential distribution occupies an important place in theory and application among the families of distributions. For this reason, distribution free confidence intervals and exceedance statistics based on these intervals need to be discussed for this class.

Consider the class of distributions

$$\mathbf{P}=\{P_\theta:P_\theta(x)=1-\exp(-\theta x),\ \ x\geq 0,\theta>0.\}$$

The parameter θ is a scale parameter for this family. Let $X_1,X_2,...,X_n$ be a random sample with d.f. $P_\theta\in\mathbf{P}_3$, so

$$f_1(X_1,X_2,...,X_n)=\sum_{i=1}^{n}a_iX_{[i]}=\sum_{i=1}^{n}a_iX_{(n-i+1)}$$

$$f_2(X_1,X_2,...,X_n)=\sum_{i=1}^{n}b_iX_{[i]}=\sum_{i=1}^{n}b_iX_{(n-i+1)},$$

and $a\prec b$.

A distribution free confidence interval for $\mathbf{P}$ is conceived by the following theorems.

Theorem 1.2. (Bairamov et al. (1999)) For the class of exponential $\mathbf{P}$, it is true that

$$P_\theta\left\{X_{n+1}\in\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)\right\}$$

$$=\quad\frac{n!}{\prod\limits_{j=1}^{n}\sum\limits_{i=1}^{j}(a_i+1)}-\frac{n!}{\prod\limits_{j=1}^{n}\sum\limits_{i=1}^{j}(b_i+1)}$$

$$=\alpha_1(a_1,a_2,...,a_n;b_1,b_2,...,b_n)=\beta_1\ ,\text{ for }\forall\theta\in\Theta,$$

and $\left(\sum_{i=1}^n a_iX_{[i]},\sum_{i=1}^n b_iX_{[i]}\right)=J(a,b,X_1,X_2,...,X_n)$ is an invariant confidence interval for the class $\wp_3$ at β_1 level.

Corollary 1.1. Let , for example, $a=(0,0,...,1)$ and $b=(1,0,...,0), a\prec b$. Then $\sum_{i=1}^n a_iX_{[i]}=X_{[n]}=X_{(1)}$, $\sum_{i=1}^n b_iX_{[i]}=X_{[1]}=X_{(n)}$ and $\alpha_1(0,0,...,1;1,0,...,0)=\frac{n-1}{n+1}$.

Theorem 1.3. (Bairamov et al.(1999)) The probability that a new set of random sample values will fall into interval $J(a,b,X_1,X_2,...,X_n)$ is

$$P_\theta\left\{X_{n+1},X_{n+2},...,X_{n+m}\in J(a,b,X_1,X_2,...,X_n)\right\}$$

$$=\quad n!\sum_{k=0}^{m}\frac{(-1)^k\binom{m}{k}}{\prod\limits_{j=1}^{n}\sum\limits_{i=1}^{j}\left\{(m-k)\,a_i+kb_i+1\right\}}$$

$$\equiv\quad\beta_m(a_1,a_2,...,a_n;b_1,b_2,...,b_n)\equiv\beta_m.$$

Corollary 1.2. Let $a=(0,0,...,1)$ and $b=(1,0,...,0), a\prec b$. Then $\sum_{i=1}^n a_iX_{[i]}=X_{[n]}=X_{(1)}$, $\sum_{i=1}^n b_iX_{[i]}=X_{[1]}=X_{(n)}$. Then one can obtain from Theorem 1.3

$$P_\theta\left\{X_{n+1},X_{n+2},...,X_{n+m}\in(X_{(1)},X_{(n)})\right\}$$

$$=\quad\frac{n!m!}{m+n}\sum_{k=0}^{m}\frac{(-1)^k}{(m-k)!(n-1+k)!}=$$

$$\frac{n!}{m+n}\sum_{k=0}^{m}\frac{(-1)^k m!k!(n-1)!}{(m-k)!k!(n-1+k)!(n-1)!}$$

$$=\frac{n!}{(m+n)(n-1)!}\sum_{k=0}^{m}(-1)^k\binom{m}{k}\binom{n-1+k}{k}^{-1}$$

$$=\frac{n!}{(m+n)(n-1)!}.\frac{n-1}{n-1+m}=\frac{n(n-1)}{(m+n)(n-1+m)}.\qquad(2)$$

Since $\wp_3 \subset \Im_c$ (2) is a special case of the following formula (see Bairamov and Petunin , 1991, Theorem 2)

$$\begin{aligned} & P_F\left\{X_{n+1}, X_{n+2}, ..., X_{n+m} \in (X_{(i)}, X_{(j)})\right\} \\ = & \frac{n!(m+j-i-1)!}{(j-i-1)!(m+n)!} \qquad \forall F \in \Im_c, \end{aligned}$$

taking $i=1$ and $j=n$. (Above in (2) we used the formula

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} \binom{a+k}{k}^{-1} = \frac{a}{a+m})$$

Theorem 1.2 and Theorem 1.3 above show that a new random sample of size m, X_{n+1}, $X_{n+2}, ..., X_{n+m}$ will have observed values that fall in the distribution free random interval $J(a, b, X_1, X_2, ..., X_n)$ and the probability of this random event is independent of the parameter θ of exponential distribution.

1.1. Distributions of exceedance statistics

Let $\phi(u_1, u_2, ..., u_n)$ be a real valued integrable n dimensional function. Consider a functional of it is defined as follows ;

$$H_F(\phi) = \int ... \int \phi(u_1, u_2, ..., u_n) dF(u_1) dF(u_2) .... dF(u_n) \ , \ F \in \mathbf{F} \ ,$$

where $\mathbf{F}$ is some class of distribution functions. The properties of functional $H_F(\phi)$ are

i) $H_F(1) = 1$

ii) $H_F(c_1\phi_1(.) + c_2\phi_2(.)) = c_1 H_F(\phi_1) + c_2 H_F(\phi_2)$.

Where $\phi_j(.)$ are distinct functions and c_j 's are real valued numbers.

Denote two random samples from two distributions $F(u)$ and $Q(u)$ as $(X_1, X_2, ..., X_n)$ and $(Y_1, Y_2, ..., Y_m)$, respectively. Let $f_1$and f_2 be the functions with properties expressed in (1). The probability of a random event

$$\begin{aligned} A_k &= \{Y_k \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\}, \\ k &= 1, 2, ...m \text{ , is} \end{aligned}$$

$$\begin{aligned} p &\equiv P(A_k) \\ &= \int ... \int \left[Q(f_2(u_1, u_2, ..., u_n)) - Q(f_1(u_1, u_2, ..., u_n))\right] dF(u_1) dF(u_2) .... dF(u_n), \end{aligned}$$

which is independent of k, as seen. If we take definition of $H_F(\phi)$ above into consideration, the required probability for the each A_k is calculated by the following;

$$P(A_k) = p = H_F\left[Q(f_2(\bar{u})) - Q(f_1(\bar{u}))\right] \equiv H_F(Q_{f_1}^{f_2}(\bar{u})),$$

where $\bar{u} = (u_1, u_2, ..., u_n)$ and $Q(f_2(\bar{u})) - Q(f_1(\bar{u})) \equiv Q_{f_1}^{f_2}(\bar{u})$. Then it is clear that

$$\xi_k = \begin{cases} 1, \text{ if random event } A_k \text{ is observed} \\ 0, \text{ if random event } A_k \text{ is not observed} \end{cases}$$

and the distribution of the exceedance statistics $\nu_m = \xi_1 + \xi_2 + ... + \xi_m$, which can take values from the set $\{0, 1, 2, ..., m\}$, can be investigate in terms of the functional H_F. Note that the r.v.'s $\xi_1, \xi_2, ..., \xi_m$ are dependent.

Theorem 1.4. (Bairamov et al. (1999)) For $k = 0, 1, 2, ..., m$ it is true that

$$P\{\nu_m = k\} = C_m^k H_F\left(\left[Q_{f_1}^{f_2}(\bar{u})\right]^k \left[1 - Q_{f_1}^{f_2}(\bar{u})\right]^{m-k}\right)$$

$$E_F\left(\left[Q_{f_1}^{f_2}(X_1, X_2, ..., X_n)\right]^k \left[1 - Q_{f_1}^{f_2}(X_1, X_2, ..., X_n)\right]^{m-k}\right)$$

where $C_m^k = \binom{m}{k} = \frac{m!}{k!(m-k)!}$, and mean and variance of ν_m are obtained as follows, respectively:

$$E(\nu_m) = mH_F(Q_{f_1}^{f_2}(\bar{u})),$$

$$\begin{aligned} var(\nu_m) &= m^2\left[H_F(Q_{f_1}^{f_2}(\bar{u}))^2 - \left(H_F(Q_{f_1}^{f_2}(\bar{u}))\right)^2\right] \\ &\quad -m\left[H_F(Q_{f_1}^{f_2}(\bar{u}))^2 - H_F(Q_{f_1}^{f_2}(\bar{u}))\right]. \end{aligned}$$

Lemma 1.1. The characteristic function for ν_m statistic is

$$\varphi_{\nu_m}(t) = H_F\left(1 + \left(e^{it} - 1\right) Q_{f_1}^{f_2}(u_1, u_2, ..., u_n)\right)^m.$$

Now let us define standardized ν_m as $\nu_m^* = \frac{\nu_m - E(\nu_m)}{\sqrt{var(\nu_m)}}$ with $E(\nu_m^*) = 0$, $var(\nu_m^*) = 1$.

Denote

$$\begin{aligned} C(x) &= P\left\{Q_{f_1}^{f_2}(X_1, X_2, ..., X_n) \leq x\right\} \\ &= P\left\{Q(f_2(X_1, X_2, ..., X_n)) - Q(f_1(X_1, X_2, ..., X_n)\ ) \leq x\right\} \end{aligned}$$

.

Theorem 1.5. (Bairamov et al.(1999)) Let f_1 and f_2 be continuous functions, F and Q are continuous d.f.'s. Then it is true that

$$\lim_{m\to\infty} \sup_{0\leq x\leq 1} \left|P\left\{\frac{\nu_m}{m} \leq x\right\} - C(x)\right| = 0.$$

The following results follow from Theorem 1.5.

Corollary 1.3. Denote $a = A(F,Q) = H_F(Q_{f_1}^{f_2}(\bar{u}))^2$ and $b = B(F,Q) = H_F(Q_{f_1}^{f_2}(\bar{u}))$. Then under conditions of Theorem 1.5 it is true

$$\lim_{m\to\infty} \sup_{-\frac{b}{a}\leq x\leq \frac{1-b}{a}} |P\{\nu_m^\star \leq x\} - C_1(x)| = 0,$$

where $C_1(x) = C(ax+b)$ and $\nu_m^\star = \frac{\nu_m - E\nu_m}{\sqrt{m^2(a-b^2)-m(a-b)}}$.

Corollary 1.4. Let $(f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))$ be the invariant confidence interval for some class of distributions $\Im$ with confidence level α_1, i.e.

$$P_F\{X_{n+1} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\} = \alpha_1 \quad \text{for any } F \in \Im.$$

Denote $\alpha_2 = P_F\{X_{n+1}, X_{n+2} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\}$, where $X_1, X_2, ..., X_n,\ X_{n+1}, X_{n+2}$ is the random sample from distribution with d.f. $F \in \Im$. Let $F = Q$ and $F \in \Im$, $X = (X_1, X_2, ..., X_n)$. Then

$$\lim_{m\to\infty} \sup_x \left| P\left\{ \frac{\nu_m - m\alpha_1}{\sqrt{m^2(\alpha_2-\alpha_1^2) - m(\alpha_2-\alpha_1)}} \leq x \right\} - C_2(x) \right| = 0,$$

where

$$C_2(x) = \begin{cases} 0, & if \quad x \leq -\frac{\alpha_1}{\sqrt{\alpha_2-\alpha_1^2}} \\ P\{F(f_2(X)) - F(f_1(X)) \leq \sqrt{\alpha_2-\alpha_1^2}x + \alpha_1\}, & if \quad x \in (-\frac{\alpha_1}{\sqrt{\alpha_2-\alpha_1^2}}, \frac{1-\alpha_1}{\sqrt{\alpha_2-\alpha_1^2}}) \\ 1, & if \quad x \geq \frac{1-\alpha_1}{\sqrt{\alpha_2-\alpha_1^2}} \end{cases}.$$

Corollary 1.5. Let **P**$=\Im_c$, where $\Im_c$ is the family of all continuous distributions. Let
$f_1(X_1, X_2, ..., X_n) = X_{(i)}, \quad f_2(X_1, X_2, ..., X_n) = X_{(j)}, 1 \leq i < j \leq n$. Given all these we can write and show that (see Bairamov, Petunin,1991)

$$H_F\left(F_{u_{(i)}}^{u_{(j)}}(u_1, u_{2,...,}u_n)\right) = P\left\{X_{n+1} \in \left(X_{(i)}, X_{(j)}\right)\right\} = \frac{j-i}{n+1} \equiv \alpha_{i,j} \text{ and}$$

$$H_F\left[\left(F_{u_{(i)}}^{u_{(j)}}(u_1, u_{2,...,}u_n)\right)^m\right] = P\left\{X_{n+1}, X_{n+2}, ..., X_{n+m} \in \left(X_{(i)}, X_{(j)}\right)\right\}$$

$$= \frac{n!(m+j-i-1)!}{(j-i-1)!(m+n)!} \equiv \alpha_{ij}^{(m)}.$$

If $i = 1$ and $j = n$ then $\alpha_{1,n} = \frac{n-1}{n+1}$, $\alpha_{1,n}^{(2)} = \frac{(n-1)n}{(n+1)(n+2)}$.

Corollary 1.6. Let $X_1, X_2, ..., X_n$ be a sample with d.f $F \in$**P**$=\Im_c$, where $\Im_c$ is the family of all continuous distributions. Let $f_1(X_1, X_2, ..., X_n) = X_{(i)}$, $f_2(X_1, X_2, ..., X_n) = X_{(j)}, 1 \leq i < j \leq n$. In this case, $C(x)$ in Theorem 1.5 has the form $C(x) = P_F\left\{Q(X_{(j)}) - Q(X_{(i)}) \leq x\right\}$. If $F = Q$ then $C(x) = P\left\{F(X_{(j)}) - F(X_{(i)}) \leq x\right\} = P\left\{W_{ij} \leq x\right\}$, where W_{ij} has beta distribution with parameter $(j-i, n-j+i+1)$ (see David, 1981). It is not difficult to see that a and b in Corollary 1.3 are:

$$a = \sqrt{\frac{(j-i)(j-i+1)}{(n+1)(n+2)} - \frac{(j-i)^2}{(n+1)^2}}, \quad b = \frac{j-i}{n+1}.$$

2. Exceedance Statistics based on minimal spacing

Let $X_1, X_2, ..., X_n$ be a sample from nonnegative continuous distribution with d.f. F, $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ be order statistics constructed by this sample. Consider the spacings $X_{1:n} - X_{0:n}$, $X_{2:n} - X_{1:n}$, $X_{3:n} - X_{2:n}, ...,$ $X_{n:n} - X_{n-1:n}$ $(X_{0:n} = 0)$. Define a random variable ν as follows: $\nu = k$ iff $X_{k:n} - X_{k-1:n} \leq X_{i:n} - X_{i-1:n}$, $i = 1, 2, .., n$. It is clear that ν is the number of a spacing having minimal length.

The following assertions are correct.

Theorem 2.1. (Bairamov (1991)) Let $F(x) = 1 - \exp(-\lambda x)$, $x \geq 0, \lambda > 0$. Then

$$P\left\{\nu = k\right\} = \frac{2(n-k+1)}{n(n+1)}, \quad k = 1, 2, ..., n.$$

Theorem 2.2. (Bairamov and Eryilmaz (2000)) Let $X_{n+1}, X_{n+2}, ..., X_{n+m}$ be the next m observations obtained independently of $X_1, X_2, ..., X_n$ from the same population with d.f. F . If $F(x) = 1 - \exp(-\lambda x)$, $x \geq 0, \lambda > 0$, then

$$P\left\{X_{n+1}, X_{n+2}, ..., X_{n+m} \in (X_{\nu-1:n}, X_{\nu:n})\right\} = \frac{(\frac{n(n+1)}{2} - 1)!m!n(m+n+1)}{(\frac{n(n+1)}{2} + m)!(m+2)} .$$

Corollary 2.1. For $m = 1$ it is true

$$P\left\{X_{n+1} \in (X_{(\nu-1)}, X_{(\nu)})\right\} = \frac{4}{3}\frac{(n+2)}{(n^2+n+2)(n+1)}.$$

Theorem 2.3. Let $X_{n+1}, X_{n+2}, ..., X_{n+m}$ be the next m observations obtained independently of $X_1, X_2, ..., X_n$ from the same population with d.f. F. If $F(x) = 1 - \exp(-\lambda x)$, $x \geq 0, \lambda > 0$, then

$$P\left\{X_{n+1}, X_{n+2}, ..., X_{n+s} \in (X_{(\nu-1)}, X_{(\nu)}), X_{n+s+1}, ..., X_{n+m} \notin (X_{(\nu-1)}, X_{(\nu)})\right\}$$

$$\begin{aligned} &= \frac{2}{n+1}\sum_{i=0}^{m-s}(-1)^i\binom{m-s}{i}\frac{n+s+i+1}{s+i+2}\left(\binom{\frac{n(n+1)}{2}+s+i}{s+i}\right)^{-1}, \\ s &= 0, 1, 2, ..., m. \end{aligned}$$

Now, define the following random variables

$$\begin{aligned} \xi_i &= \begin{cases} 1 & \text{if} \quad X_{n+i} \in (X_{\nu-1:n}, X_{\nu:n}) \\ 0 & \text{if} \quad X_{n+i} \notin (X_{\nu-1:n}, X_{\nu:n}) \end{cases}, \\ i &= 1, 2, ..., m, \ \ X_{(0)} = 0 \text{ and the exceedance statistic } S_m = \sum_{i=1}^{m}\xi_i. \end{aligned}$$

It is clear that the random variables $\xi_1, \xi_2, ..., \xi_n$ are dependent.

The following statement is valid.

Theorem 2.4. (Bairamov and Eryilmaz (2000)) Let $F(x) = 1 - \exp(-\lambda x)$, $x \geq 0, \lambda > 0$. Then the distribution of random variable S_m is

$$P\{S_m = s\}$$

$$= \binom{m}{s}\frac{2}{n+1}\sum_{i=0}^{m-s}(-1)^i\binom{m-s}{i}\frac{n+s+i+1}{s+i+2}\left(\binom{\frac{n(n+1)}{2}+s+i}{s+i}\right)^{-1},$$

$s = 0, 1, 2, ..., m.$

2.1. Asymptotic distribution of S_m for any continuous F

Theorem 2.5 The asymptotic distribution of $\frac{S_m}{m}$ for large m is

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

One can observe that Theorem 2.5 can be extended as follows: let $X_1, X_2, ..., X_n$ be a sample with continuous d.f. F, $Y_1, Y_2, ..., Y_n$ be a sample with continuous d.f. G. Define the following random variables

$$\bar{\xi}_i = \begin{cases} 1 & \text{if} \quad Y_i \in (X_{\nu-1:n}, X_{\nu:n}) \\ 0 & \text{if} \quad Y_i \notin (X_{\nu-1:n}, X_{\nu:n}) \end{cases},$$

$$i = 1, 2, ..., m;\ X_{(0)} = 0 \text{ and } \bar{S}_m = \sum_{i=1}^{m} \bar{\xi}_i.$$

Theorem 2.5A. The asymptotic distribution of $\frac{\bar{S}_m}{m}$ for large m is

$$P\{G(X_{\nu:n}) - G(X_{\nu-1:n}) \le x\}$$

i.e. it is true that

$$\lim_{m\to\infty} \sup_{0\le x\le 1} \left| P\left\{\frac{\bar{S}_m}{m} \le x\right\} - P\{G(X_{\nu:n}) - G(X_{\nu-1:n}) \le x\}\right| = 0.$$

Lemma. Let X_1, X_2, X_3 be iid random variables with cıontinuous d.f. F. Let $X_{\nu:3} - X_{\nu-1:3} \le X_{i:3} - X_{i-1:3}$, $i = 2, 3$. The pdf of random variable $F(X_{\nu:3}) -$

$F(X_{\nu-1:3})$ is

$$f^*(t) = 6\int_t^1 [1 - F(2F^{-1}(v) - F^{-1}(v-t)) + F(2F^{-1}(v-t) - F^{-1}(v))]dv$$

$$0 < t < 1.$$

Example. Let X_1, X_2, X_3 be a sample from uniform distribution. Then

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

and the distribution function is

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

2.3. Exceedance models of Wesolowski and Ahsanullah

Wesolowoski and Ahsanullah (1998) have introduced several interesting exceedance models and investigated the distributional properties of these statistics.

They also have discussed the identification of equiedistribution of observations and the threshold. We mention here only few of their results.

Wesolowoski and Ahsanullah (1998) consider the number of $Y's$ not exceeding the level X and counted to the moment of the n^{th} exceedance. Precisely, for any integer $n \geq 1$ the exceedance statistics is defined as

$$R_n = \min\{j \geq 0 : S_{n+j-1} = j, Y_{n+j} > X\} = \min\{j \geq 0 : Y_{j+1:n+j} > X\}.$$

The exact and assymptotic distributions of R_n is given in the following

Theorem A. Assume that $P\{Q(X) < 1\} > 0$. For any integer $n \geq 1$

$$P\{R_n = j\} = \binom{n+j-1}{n-1} E(Q^j(X)(1-Q(X)), j = 0, 1, ...$$

and

$$E(R_n) = nE\left(\frac{Q(X)}{1-Q(X)} I_{(0,1)}(Q(X))\right)$$

$$\begin{aligned} Var(R_n) &= nE\left(\frac{Q(X)}{(1-Q(X))^2} I_{(0,1)}(Q(X))\right) \\ &+n^2 Var\left(\frac{Q(X)}{(1-Q(X))^2} I_{(0,1)}(Q(X))\right). \end{aligned}$$

Theorem B. Assume that $Q(X) < 1$ a.s. Then for $n \to \infty$

$$\frac{1}{n} R_n \xrightarrow{d} \frac{Q(X)}{1-Q(X)}.$$

3. Exceedance statistics based on record values

Let $X_1, X_2, ..., X_n, ...$ be a sequence of independent and identically distributed random variables with continuous distribution function F, $X_{u(1)}, X_{u(2)}, ...$ be a corresponding sequence of record values.

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

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

It is known that the distribution function (Ahsanullah (1995), Arnold, Balakrishnan, Nagaraja (1998)) of n th record value is

$$F_n(x) = P\left\{X_{u(n)} \leq x\right\} = \int\limits_{-\infty}^{x} \frac{R^{n-1}(u)}{(n-1)!} dF(u) \qquad , -\infty < x < \infty.$$

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

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

Suppose that $X_{u(n)+1}, X_{U(n)+2}, ..., X_{U(n)+m}$ be the m observations that come after $X_{u(n)}$.

For $i = 1, 2, ..., m$ and $r < s$,

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

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

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

Denote $\nu_m = \sum\limits_{i=1}^{m} \xi_i$, the number of observations $X_{u(n)+1}, X_{u(n)+2}, ..., X_{u(n)+m}$ falling into interval $(X_{u(r)}, X_{u(s)})$. It is clear that the r.v.'s $\xi_1, \xi_2, ..., \xi_m$ are dependent.

Theorem 3.1. For $r = k-1$ and $s = k$ ($k = 2, 3, ...$), it is true that

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

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

(see Bairamov (1997)).

3.1. Assymptotic distributions of exceedance statistics in a record threshold model

Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. r.v.'s with continuous d.f. F , $X_{u(1)}, X_{u(2)}, ...$ is a sequence of record values. Let $X_1', X_2', ..., X_m'$ be i.i.d. random variables with continuous d.f. G. Define the following random variables

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

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 $X_{u(s)}$.

Lemma 3.1. It is true that

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

The proof of this lemma is based on the Glivenko-Cantelli theorem.

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

Corollary 2. Let $F = G$. Then

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

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

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

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 $r = k-1, s = k$ $(k = 2, ..., n)$. Then the expected value and the variance of $S_m(r, s)$ is:

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

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

As a consequence of Lemma 3.1, we have the following:

Theorem 3.2. Let $F = G$. For $r = k-1$ and $s = k$ $(k = 2, 3, ..., n)$

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

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

$$D_k(x) \equiv P\left\{ U_{k-1,k} \le x \right\} = \frac{1}{(k-1)!} \int\limits_0^x \left[\ln \frac{1}{u} \right]^{k-1} du \ , \quad 0 < x < 1.$$

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

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

where

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

(see Bairamov and Eryılmaz (2001))

4. Exceedance models in multivariate FGM distributions

Let us define a n variate FGM random vector $(X_1, X_2, ..., X_n)$ by the univariate marginals $F_i \sim X_i$, $i = 1, 2, ..., n$ and a real number α_n such that the joint distribution of $X_1, X_2, ..., X_n$ is given by the Farlie-Gumbel-Morgenstern (FGM) distribution

$$H_{1,2,...,n}(\mathbf{x}) = \prod_{i=1}^{n} F_i(x_i) \left\{ 1 + \alpha_n \sum_{1\le j<l\le n} \bar{F}_j(x_j)\bar{F}_l(x_l) \right\}, \tag{3}$$

where $\mathbf{x} = (x_1, x_2, ..., x_n) \in R^n$, $\bar{F}_j(x_j) = 1 - F_j(x_j)$. For a given n the real number α_n is admissible if

$$1 + \alpha_n \sum_{1\le j<l\le n} \varepsilon_j \varepsilon_l \ge 0$$

hold for all $\varepsilon_j = \pm 1$ (as a consequence each parameter must satisfy $|\alpha| \leq 1$).

We call the random variables whose joint distribution is defined by (3) a simple- FGM (or s-FGM) random variables. (If it will not confuse the readers we will also use the term "finite s-FGM sequence")

It follows from the Lemma 4.1 below that if n tends to infinity, i.e. if we consider infinite sequence of random variables whose finite dimensional distributions are given in (3) then we have only a sequence of independent random variables. We will deal in this paper with finite FGM sequences.

Lemma 4.1. The admissible range of α_n, $n > 1$ allowing (3) to be a n variate distribution function is

$$-\frac{1}{\binom{n}{2}} \leq \alpha_n \leq \frac{1}{\left[\frac{n}{2}\right]},$$

where $[x]$ denotes the integer part of the real number x.

Consider a random variable X with the continuous distribution function F and s-FGM random variables $Y_1, Y_2, ..., Y_m$ independent of X and fitting to the model (3) with equal marginal distributions $G(x_j)$ and pdf $g(x_j)$, $-\infty \leq x_j \leq \infty$, $j = 1, 2, ...m$. The joint pdf of $(Y_1, Y_2, ..., Y_m)$ is given as follows:

$$h(x_1, x_2, ..., x_m) = \prod_{i=1}^{m} g(x_i) \left\{ 1 + \alpha_m \sum_{1 \leq j < l \leq m} (1 - 2G(x_j))(1 - 2G(x_l)) \right\},$$

$$-\infty \leq x_1, x_2, ..., x_m \leq \infty$$

with α_m satisfying

$$-\frac{1}{\binom{m}{2}} \leq \alpha_m \leq \frac{1}{\left[\frac{m}{2}\right]} \tag{4}$$

Definition 4.1. For any integer $m \geq 1$

$$S_m = \# \{k \leq m : Y_k \leq X\}$$

denotes the number of Y's falling below the random threshold X.

The exact distribution of S_m is given in the following theorem.

Theorem 4.1. (Bairamov and Eryilmaz (2004)) For any integer $m \geq 1$ and real number α_m satisfying (4)

$$P\{S_m = k\}$$

$$= \binom{m}{k} \left[E(G^k(X)\bar{G}^{m-k}(X)) + \alpha_m \left(\frac{k(k-1)}{2} E(G^k(X)\bar{G}^{m-k+2}(X)) \right.\right.$$

$$-k(m-k)E(G^{k+1}(X)\bar{G}^{m-k+1}(X))+$$
$$+\frac{(m-k)(m-k-1)}{2}E(G^{k+2}(X)\bar{G}^{m-k}(X))\Big)\Big]$$
$$k=0,1,...,m,$$

where $\bar{G}(x)=1-G(x)$.

Corollary 4.1. Let $F=G$, then

$$P\{S_m=k\}$$

$$=\binom{m}{k}\Big[B(k+1,m-k+1)+\alpha_m\Big(\frac{k(k-1)}{2}B(k+1,m-k+3)$$
$$-k(m-k)B(k+2,m-k+2)+$$
$$+\frac{(m-k)(m-k-1)}{2}B(k+3,m-k+1)\Big)\Big] \qquad (5)$$
$$k=0,1,...,m, \qquad -\frac{1}{\binom{m}{2}}\leq\alpha_m\leq\frac{1}{\left[\frac{m}{2}\right]}.$$

where $B(a,b)$ is a beta function.

Remark 4.1. Since $\binom{m}{m-k}=\binom{m}{k}$ and $B(a,b)$ is a symmetric function of its arguments it is not difficult to observe from (5) that $P\{S_m=k\}=P\{S_m=m-k\}$.

Let $X_1,X_2,...,X_n$ be the finite s-FGM sequence having marginal d.f. F and $Y_1,Y_2,...,Y_m,...$ be a sequence of i.i.d. random variables with d.f. G. Let $X_{1:n}\leq X_{2:n}\leq ...\leq X_{n:n}$ be the order statistics of $X_1,X_2,...,X_n$.

Definition 4.2. For any integer $m\geq 1$ and $1\leq r\leq n$

$$\nu_m=\#\{k\leq m:Y_k\leq X_{r:n}\}$$

denotes the number of Y's falling below the random threshold $X_{r:n}$.

The exact distribution of ν_m is given in the following theorem.

Theorem 4.2. If $H_0:F=G$ is true then for any integer $m\geq 1$ and $1\leq r\leq n$

$$P\{\nu_m=k\}$$

$$= \binom{m}{k} \sum_{s=r}^{n} (-1)^{s-r} \binom{s-1}{r-1} \binom{n}{s} [sB(s+k, m-k+1)$$

$$+\alpha_n \left(\frac{s^2(s-1)}{2} B(s+k, m-k+3) - s(s-1)B(s+k+1, m-k+2) \right) \Big]$$

$$k = 0, 1, ..., m \quad , \quad -\frac{1}{\binom{n}{2}} \leq \alpha_n \leq \frac{1}{\left[\frac{n}{2}\right]} .$$

Remark 4.2. Let in Theorem 4.1 $\alpha_m = 0$ i.e. $Y_1, Y_2, ..., Y_m$ are i.i.d. random variables. Then from Theorem 4.1 one obtains

$$P\{S_m = k\} = \binom{m}{k} E(G^k(X)\bar{G}^{m-k}(X))$$

which coincides with the Theorem 1 of Wesolowski and Ahsanullah (1998). From Theorem 4.2 in this case one has

$$P\{\nu_m = k\} = \frac{\binom{r+k-1}{r-1}\binom{m+n-k-r}{n-r}}{\binom{m+n}{n}}$$

which coincides with Corollary 1 (b) of Wesolowski and Ahsanullah (1998) (see also Katzenbeisser (1985), (1986), Matveychuk and Petunin (1990), Johnson and Kotz (1991), (1994)).

5. Exceedance statistics for arbitrary distributions

Let X be a random variable defined on a probability space $\{\Omega, \Im, P\}$ with c.d.f. $F(x) = P\{X \leq x\}$. Throughout this paper we will assume that F is an arbitrary distribution, i.e. F may contain a discrete, absolutely continuous and singular components. Let $M = \{a_1, a_2, ..., a_l\}$, $(a_1 < a_2 < ... < a_l)$ be the set of atoms of the distribution. The following lemma will be useful for our investigations.

Lemma 5.1. Let $A \in \Im$ and $E\{A \mid x\}$ exist for all $x \in R$. Then

$$P\{A\} = \sum_{k=0}^{l} \int_{a_k}^{a_{k+1}-0} P\{A \mid X = x\} \, dF(x) + \sum_{k=1}^{l} P\{A \mid X = a_k\} P\{X = a_k\}, \tag{6}$$

where $a_0 = -\infty, a_{l+1} = \infty$.

The proof is a direct application of the total probability rule.

It is clear that if we have only one atom, say a then (6) becomes

$$P\{A\} = \int_{-\infty}^{a-0} P\{A \mid X = x\} \, dF(x) + \int_{a}^{\infty} P\{A \mid X = x\} \, dF(x)$$

$$+P\{A \mid X=a\}\,P\{X=a\}. \tag{7}$$

Now let $X_1, X_2, ..., X_n$ be a sample from distribution with the c.d.f. F, and $X_{n+1}, X_{n+2}, ..., X_{n+m}$ be the another sample from the same distribution independent of the first. Denote by $X_{(1)}, X_{(2)}, ..., X_{(n)}$ the order statistics of $X_1, X_2, ..., X_n$. Consider the random variables $\xi_1(r), \xi_2(r), ..., \xi_m(r)$ defined as follows:

$$\xi_i(r)=\begin{cases} 1, & if \quad X_{n+i}<X_{(r)} \\ 0 & if \quad X_{n+i}\geq X_{(r)} \end{cases}, i=1,2,...,m;\ 1\leq r\leq n.$$

Finally define as above

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

Evidently $S_m(r)$ is the number of observations $X_{n+1},\ X_{n+2}, ...,\ X_{n+m}$ falling in to the interval $(-\infty, X_{(r)})$. It is well known that the distribution function of $X_{(r)}$ is

$$F_r(x)=P\left\{X_{(r)}\leq x\right\}=\sum_{i=r}^{n}\binom{n}{i}F^i(x)(1-F(x))^{n-i}=I_{F(x)}(r,n-r+1),$$

where $I_p(c,d)=\int\limits_0^p t^{c-1}(1-t)^{d-1}dt/B(c,d)$ is the uncompleted beta function (see e.g. David (1981))

Theorem 5.1. Using the notation above:

$$P\{S_m(r)=k\}=\binom{m}{k}\left\{\frac{1}{B(r,n-r+1)}\sum_{j=0}^{l}\int\limits_{F(a_j)}^{F(a_{j+1}-0)}t^{k+r-1}(1-t)^{m+n-k+r}dt\right.$$

$$\left.+\left(F_r(a)-F_r(a-0)\right)\sum_{j=1}^{l}F^k(a_j-0)\left((F(a_j)-F(a_j-0)\right)^{m-k}\right\},$$

$$k=0,1,...,m.$$

Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables with distribution F of arbitrary type and let D be the set of points of discontinuity of F. Throughout $X_{1:n}\leq X_{2:n}\leq\cdots\leq X_{n:n}$ will represent the order statistics from a sample of size n and $F_{r:n}$ will be the distribution of $X_{r:n}$. For a second sample, $X_{n+1},\cdots,X_{n+m}$, define the exceedance statistic, $S_m(r)$, to be the number of X_{n+i} that fall strictly below $X_{r:n}$. (We should remark that one may modify our results if exceedance is defined as $X_{n+i}\leq X_{r:n}$, however we will not duplicate such results.) The sampling distribution of $S_m(r)$ is as follows.

Theorem 5.2. (Bairamov and Khan (2004)) For $k = 0, 1, \cdots, m$, we have

$$\frac{P(S_m(r)=k)}{\binom{m}{k}} = \frac{r\binom{n}{r}}{(m+n)\binom{m+n-1}{k+r-1}}\left(1-\sum_{d\in D}(\Delta F_{k+r:m+n})_d\right) + \sum_{d\in D}(F(d^-))^k\,(1-F(d^-))^{m-k}(\Delta F_{r:n})_d,$$

where $(\Delta F_{r:n})_d = F_{r:n}(d) - F_{r:n}(d^-)$, and $\sum_{d\in D}$ stands for summing over all points of discontinuity of F.

When D is empty, one gets the classical result of Gumbel and Schelling(1950), see also Gumbel (1954), Epstein (1954), and Sarkadi (1954). When D is a finite set, one gets the recent result of Bairamov and Kotz (2001).

Our next result deals with the weak convergence of the statistic $S_m(r)/m$. When F is a continuous distribution, the weak convergence of $S_m(r)/m$ to a beta distribution was proved by Matveychuck and Petunin (1990, (1991). The following theorem not only provides the weak limit of $S_m(r)/m$ for an arbitrary F but also gives its rate of convergence.

Theorem 5.3. For any distribution F, and any $x \in (0,1)$, the rate of weak convergence is

$$\left|P(S_m(r)\le mx) - \widehat{L_{r,n}}(x)\right| = O(m^{-1/2}), \text{ where}$$

$$\widehat{L_{r,n}}(x) :=$$

$$= \begin{cases} B_{r,n-r+1}(F(d)) & \text{if} \quad \begin{array}{c} x\in(F(d^-),F(d)] \\ \text{for some } d\in D, \end{array} \\ \frac{1}{2}[B_{r,n-r+1}(F(d)) + B_{r,n-r+1}(F(d^-))] & \text{if} \quad \begin{array}{c} x = F(d^-) \\ \text{for some } d\in D, \end{array} \\ B_{r,n-r+1}(x) & \text{otherwise} \end{cases},$$

where $B_{a,b}(x)$ is the beta distribution with parameters a, b. (Of course, the weak limit is the right continuous version of $\widehat{L_{r,n}}$.)

In particular, when F is a continuous distribution, we get

$$|P(S_m(r)\le mx) - B_{r,n-r+1}(x)| = O(m^{-1/2}).$$

References

Ahsanullah, M. (1995) *Records.* Nova Science Publishers. Huntigton, New York.

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

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

Bairamov I.G. (1991) Distributions related with differences of addjacent order statistics. *Transactions of Academy of Sciences of Azerbaijan. Series of Physical-Technical and Mathematical Sciences.* Vol. **12**. No 1, p. 42-44.

Bairamov, I.G. (1997) Some distribution free properties of statistics based on record values and characterizations of the distributions trough a record. *J. Appl.Statist. Sci.*, **5**, 1, 17-25.

Bairamov I.G., Gebizlioglu O.L. and Kaya M.F. (1999) Distributional properties of statistics based on invariant confidence intervals. *Istatistik. Journal of the Turkish Statistical Association.* Vol. **2**. No 1, 15-32.

Bairamov, I.G. and Eryilmaz, S. (2000) Distributional properties of statistics based on minimal spacing and record exceedance statistics. *Journal of Statistical. Planning and . Inference,* **90**, 21-33.

Bairamov, I.G. and Kotz, S. (2001) On distributions of exceedances associated with order statistics and record values for arbitrary distributions. *Statistical . Papers.* Vol. **42**, Issue 2, 171-185. D

Bairamov I.G. and Eryilmaz S.N. (2001) On properties of statistics connected with minimal spcing and record exceedances. In *Applied Statistical Scinec* V, Eds. M. Ahsanullah, J. Kennyon and S.K. Sarkar, 245-254.

Bairamov I.G. and Eryilmaz S.N. (2004) Characterization of symmetry and exceedance models in multivariate FGM distributions. *Journal of Applied Statistical Science.* Volume **13**, Number 2, 87-99.

Bairamov I. and Khan M. K. (2004) On exceedances of record and order statistics. Preprint in Elsevier Science: MPS: Statistics/0401010, Statistics, 22 January, 2004.

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

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

Gumbel, E.J. and Schelling, H, von (1950) The distribution of the number of exceedances. *Annals of Mathematical Statistics*, **21**, 247-262.

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

Johnson, N. and Kotz, S. (1994) Further comments on Matveychuk and Petunin's generalized Bernoulli model, and nonparametric tests of homogeneity. *Journal of Statistical Planning and Inference*, **41**, 61-72.

Katzenbeisser, W. (1985) The distribution of two sample location exceedance test statistics under Lehmann alternatives. *Statistishe 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.

Marshal A.W., Olkin I. (1979). *Inequalities. Theory of majorization and its applications.* Academic Press.

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

Robbins, H. (1944) On distribution free tolerance limits in random sampling. *Annals of Mathematical Statistics.*, Vol. **15**, P.70-79

Sarkadi, K. (1957) On the distribution of the number of exceedances. *Annals of Mathematical Statistics*, **28**, 1021-1022.

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

Wesolowski, J. and Ahsanullah, M. (1998) Distributional properties of exceedance statistics. *Annals of the Institute of . Statistical Mathematics.*, **50**, 543-565.

Wilks S.S. (1941) Statistical prediction with special references to the problem of tolerance limits, *Annals of Mathematical Statistics*, **13**, 400-409.