Journal of Statistical Planning and Inference 90 (2000) 21–33

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Distributional properties of statistics based on minimal spacing and record exceedance statistics

Ismihan G. Bairamov *, Serkan N. Eryılmaz

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

Received 17 June 1999; received in revised form 10 December 1999; accepted 15 February 2000

Abstract

Exact and asymptotic distributions of some statistics based on spacing having minimal length are investigated. The behaviour of a sequence of independent identically distributed random variables with respect to a record threshold is studied. © 2000 Elsevier Science B.V. All rights reserved.

MSC: primary 62F20; 62G30; secondary 62E20

Keywords: Order statistics; Spacing; Record values; Exceedance statistics

1. Introduction

This work is an attempt to study the properties of a sequence of observations that fall in some random interval based on order statistics and record values. As such, it relates to the theory of tolerance limits and exceedance statistics. The latter obviously examines a sequence of subsequent random samples that fall in an interval based on an original sample, with all the samples coming from an identical distribution. Our study is also akin 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). Needless to add that tolerance limits and exceedance statistics are widely discussed by Wilks (1941,1942), Robbins (1944), Gumbell and von Shelling (1950), Epstein (1954), Sarkadi (1957) and Siddiqui (1970) along with others. Further detailed discussion can be found in David (1981), Leadbetter et al. (1983) and Johnson et al. (1992). A recent interest in the subject lies in the work of Wesolowsky and Ahsanullah (1999).

* Corresponding author.

E-mail address: Ismihan.Bayramov@science.ankara.edu.tr (I.G. Bairamov).

PII: S0378-3758(00)00107-5

Invariant confidence intervals containing the main distributed mass were introduced in 1990 (cf., 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.) $F\in\Im$, $\Im$ is some class of d.f.'s. $f_1(.)$ and $f_2(.)$ are assumed to be two Borel functions satisfying $f_1(x_1,x_2,\dots,x_n)\leqslant f_2(x_1,x_2,\dots,x_n)$, $\forall(x_1,x_2,\dots,x_n)\in\mathbb{R}^n$. 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 $\Im$, if $\exists\beta\in(0,1)$ such that

$$P\{X_{n+1}\in(f_1(X_1,X_2,\dots,X_n),f_2(X_1,X_2,\dots,X_n))\}=\beta,\quad \forall F\in\Im.$$

The quantity β is the same for all $F\in\Im$ and is called a confidence level of an invariant interval. Let $\Im=\Im_c$ be the class of all continuous d.f.'s. It is known that under some natural conditions any invariant confidence interval for the class $\Im_c$ can be generated only by the order statistics, so that $f_1(X_1,X_2,\dots,X_n)=X_{(i)}$, $f_2(X_1,X_2,\dots,X_n)=X_{(j)}$, $1\leqslant i<j\leqslant n$, where $X_{(1)}<X_{(2)}<\cdots<X_{(n)}$ are order statistics generated by $X_1,X_2,\dots,X_n$, which is defined with probability 1 (cf., Bairamov and Petunin, 1991), and

$$P\{X_{n+1}\in(X_{(i)},X_{(j)})\}=\frac{j-i}{n+1}. \tag{1.1}$$

Let $Y_1,Y_2,\dots,Y_m$ be a sample of size m with d.f. G. Throughout the paper we are considering $X_1,X_2,\dots,X_n$ and $Y_1,Y_2,\dots,Y_m$ as an independent observation on an independent pair of random variables (r.v.'s) X and Y, respectively. Let $Y_{(1)}<Y_{(2)}<\cdots<Y_{(m)}$ be the order statistics generated by $Y_1,Y_2,\dots,Y_m$. It is clear that, under hypothesis H_0: $F=G$, (1.1) may be rewritten as follows:

$$P\{Y_k\in(X_{(i)},X_{(j)})\}=\frac{j-i}{n+1},\quad 1\leqslant i<j\leqslant n,\ \ k=1,2,\dots,m.$$

Let S_m^{rs} denotes the number of observations $Y_1,Y_2,\dots,Y_m$ falling into interval $(X_{(r)},X_{(s)})$. If $F=G$ then S_m^{rs}/m has the limiting distribution $P\{W_{rs}\leqslant x\}$, where $W_{rs}=F(X_{(s)})-F(X_{(r)})$ (cf., Bairamov, 1997). W_{rs} has the probability density function (p.d.f.) (cf., David, 1981)

$$f(w_{rs})=\begin{cases}\frac{1}{B(s-r,n-s+r+1)}w_{rs}^{s-r-1}(1-w_{rs})^{n-s+r} & \text{if } 0\leqslant w_{rs}\leqslant 1,\\ 0 & \text{otherwise.}\end{cases}$$

When F is not necessarily equal to G the following assertion is also valid:

Theorem 1. *For any r and s satisfying $1\leqslant r<s\leqslant n$, it is true that*

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

Theorem 1 is extended for any random interval $(f_1(X_1,X_2,\dots,X_n),\ f_2(X_1,X_2,\dots,X_n))$ and in this case the limiting distribution of S_m^{rs} (cf., Bairamov, 1997; Bairamov and Gebizlioglu, 1998) is equal to $P\{G(f_2(X_1,X_2,\dots,X_n))-G(f_1(X_1,X_2,\dots,X_n))\}\leqslant x$.

Let H_0 is true, i.e. $F=G$. Then from Theorem 1 it follows that for $2 \leqslant i \leqslant n$

$$\lim_{n\to\infty} \lim_{m\to\infty} \sup_{-\infty<x<\infty} \left| P\left\{ \frac{nS_m^{i-1,i}}{m} \leqslant x \right\} - F_0(x) \right| = 0,$$

where $F_0(x) = 1 - \mathrm{e}^{-x}$, $x \geqslant 0$.

This paper has a two-fold purpose. First, we investigate the distributional properties of some statistics based on the so-called minimal spacing, i.e. the difference of adjacent order statistics having minimal length. Next, we present some results on exact and asymptotic distributions of some useful statistics in a record threshold model.

Let $\{X_n\}_{n\geqslant 1}$ be a sequence of i.i.d. random variables with continuous d.f. $F(x)$. Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the almost sure (a.s.) defined order statistics generated by a $X_1, X_2, \ldots, X_n$. The main result of Section 2 of the paper (Theorem 3) states that if $F(x) = 1 - \exp(-\lambda x)$, $x \geqslant 0, \lambda > 0$, then, for $s = 0, 1, \ldots, m$,

$$\begin{aligned} &P\{X_{n+1}, X_{n+2}, \ldots, X_{n+s} \in (X_{(v-1)}, X_{(v)}), X_{n+s+1}, \ldots, X_{n+m} \notin (X_{(v-1)}, X_{(v)})\} \\ &\quad = \frac{2}{n+1} \sum_{i=0}^{m-s} (-1)^i \binom{m-s}{i} \frac{n+s+i+1}{s+i+2} \left(\binom{\frac{n(n+1)}{2} + s + i}{s+i} \right)^{-1}, \\ &\quad s = 0, 1, 2, \ldots, m, \end{aligned} \tag{1.2}$$

where v is the index of the a.s. unique spacing $(X_{(k-1)}, X_{(k)})$, $k=1, \ldots, n$, having minimal length. Note that $X_{(0)}=0$. If $s=0/m$, the left-hand side (l.h.s.) term in (1.2) is interpreted as $P\{X_{n+1}, X_{n+2}, \ldots, X_{n+m} \in / \notin (X_{(v-1)}, X_{(v)})\}$. Let S_m be the number of k, $k \in \{n+1, \ldots, n+m\}$, such that $X_k \in (X_{(v-1)}, X_{(v)})$. Denote by $A(m,n,s)$ the right-hand side (r.h.s.) in (1.2). An almost straightforward consequence of (1.2) is that (Theorem 4)

$$P\{S_m = s\} = \binom{m}{s} A(m,n,s), \quad s = 0, 1, 2, \ldots, m. \tag{1.3}$$

In Section 3 we consider a general continuous d.f. F and prove that (Theorem 5)

$$\lim_{m\to\infty} \sup_{0\leqslant x\leqslant 1} \left| P\left\{ \frac{S_m}{m} \leqslant x \right\} - P\{F(X_{(v)}) - F(X_{(v-1)}) \leqslant x\} \right| = 0$$

$(X_{(0)} = -\infty)$.

In Section 4, a result similar to (1.3) is established, considering the sequence of record values $\{X_{u(n)}\}_{n\geqslant 1}$ generated by a sequence of i.i.d. r.v.'s $\{X_n\}_{n\geqslant 1}$ with a general continuous distribution function F, where

$$u(1) = 1 \quad \text{and} \quad u(n) = \min\{j\colon j > u(n-1), X_j > X_{u(n-1)}\}, \quad n > 1.$$

Let $1 \leqslant k \leqslant n$ be fixed and v_m be the number of $l, l \in \{u(n)+1, \ldots, u(n)+m\}$ such that $X_l \in (X_{u(k-1)}, X_{u(k)})$.

The subsequent result (Theorem 6) consists in the formulation: for $r = k-1$ and $s = k$ $(k = 2, \ldots, n)$, it is true that

$$P\{v_m = j\} = \binom{m}{j} \sum_{i=0}^{m-j} (-1)^i \binom{m-j}{i} \frac{1}{(j+i+1)^k}, \quad k = 2, \ldots, n.$$

The last results of this section (Theorem 7) consist essentially in establishing the following limit distributions for v_m/m and $(v_m - E(v_m))/\sigma(v_m)$:

$$\lim_{m\to\infty} \sup_{0\leqslant x\leqslant 1} \left| P\left\{\frac{v_m}{m} \leqslant x\right\} - \frac{1}{(k-1)!}\int_0^x \left[\ln\frac{1}{u}\right]^{k-1} \mathrm{d}u \right| = 0$$

and

$$\lim_{m\to\infty} P\left\{\frac{v_m - E(v_m)}{\sigma(v_m)} \leqslant x\right\} = \frac{1}{(k-1)!}\int_0^{ax+b} \left[\ln\frac{1}{u}\right]^{k-1} \mathrm{d}u,$$

where

$$Ev_m = m/2^k, \quad \mathrm{Var}(v_m) = m(1/2^k - 1/3^k) + m^2(1/3^k - 1/2^{2k}),$$

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

From an application point of view, the findings of this paper can be used to test goodness of fit or homogeneity of data.

2. Statistics based on minimal spacing

Let $X_1, X_2, \ldots, X_n$ be a sample from nonnegative continuous distribution with d.f. F. Also, let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the order statistics constructed on the basis of this sample. Consider the spacings $X_{(1)} - X_{(0)}$, $X_{(2)} - X_{(1)}$, $X_{(3)} - X_{(2)}, \ldots, X_{(n)} - X_{(n-1)}$ with $X_{(0)} = 0$. Define a random variable v as follows: $v = k$ iff $X_{(k)} - X_{(k-1)} \leqslant X_{(i)} - X_{(i-1)}$, $i = 1, 2, \ldots, n$. It is clear that v is the index of a spacing having minimal length.

It is well known that if the underlying distribution is exponential with parameter λ, then $X_{(k)} - X_{(k-1)}$, $k = 1, \ldots, n$, $(X_{(0)} = 0)$ are independent, $X_{(k)} - X_{(k-1)}$ are exponential with parameter $(n-k+1)\lambda$. Furthermore, vector of spacings $(X_{(1)}, X_{(2)} - X_{(1)}, X_{(3)} - X_{(2)}, \ldots, X_{(n)} - X_{(n-1)})$ does not depend on order statistics $X_{(1)}, X_{(2)}, X_{(3)}, \ldots, X_{(n-1)}$ (cf., e.g., Arnold et al., 1992). The following theorems can be concluded from these results.

Theorem 2. *Let* $F(x) = 1 - \exp(-\lambda x)$, $x \geqslant 0, \lambda > 0$. *Then*

$$P\{v = k\} = \frac{2(n-k+1)}{n(n+1)}, \quad k = 1, 2, \ldots, n.$$

Theorem 3. *Let* $X_{n+1}, X_{n+2}, \ldots, X_{n+m}$ *be the next* m *observations obtained independently of* $X_1, X_2, \ldots, X_n$ *from the same population with d.f.* F. *If* $F(x) = 1 - \exp(-\lambda x)$, $x \geqslant 0, \lambda > 0$, *then*

$$\begin{aligned} &P\{X_{n+1}, X_{n+2}, \ldots, X_{n+s} \in (X_{(v-1)}, X_{(v)}), X_{n+s+1}, \ldots, X_{n+m} \notin (X_{(v-1)}, X_{(v)})\} \\ &\quad = \frac{2}{n+1}\sum_{i=0}^{m-s}(-1)^i \binom{m-s}{i} \frac{n+s+i+1}{s+i+2}\left(\binom{\frac{n(n+1)}{2}+s+i}{s+i}\right)^{-1}, \\ &\quad s = 0, 1, 2, \ldots, m. \end{aligned} \tag{2.1}$$

Proof. One observes that the probability of the event

$$\{X_{n+1,}X_{n+2},\dots,X_{n+s}\in(X_{(v-1)},X_{(v)}),X_{n+s+1},\dots,X_{n+m}\notin(X_{(v-1)},X_{(v)})\}$$

can be calculated as follows:

$$P\{X_{n+1},\dots,X_{n+s}\in(X_{(v-1)},X_{(v)}),X_{n+s+1},\dots,X_{n+m}\notin(X_{(v-1)},X_{(v)})\}$$
$$=\sum_{k=1}^{n}P\{X_{n+1},\dots,X_{n+s}\in(X_{(k-1)},X_{(k)}),\ X_{n+s+1},\dots,X_{n+m}\notin(X_{(k-1)},X_{(k)}),v=k\}. \tag{2.2}$$

Consider the summand probability at the r.h.s. of (2.2). By using the independence of spacings this probability can be calculated as follows:

$$P\{X_{n+1},\dots,X_{n+s}\in(X_{(k-1)},X_{(k)}),\quad X_{n+s+1},\dots,X_{n+m}\notin(X_{(k-1)},X_{(k)}),v=k\}$$
$$=P\{X_{n+1},\dots,X_{n+s}\in(X_{(k-1)},X_{(k)}),X_{n+s+1},\dots,X_{n+m}\notin(X_{(k-1)},X_{(k)}),$$
$$X_{(k)}-X_{(k-1)}\leqslant X_{(1)},\dots,X_{(k)}-X_{(k-1)}\leqslant X_{(k-1)}-X_{(k-2)},$$
$$X_{(k)}-X_{(k-1)}\leqslant X_{(k+1)}-X_{(k)},\dots,X_{(k)}-X_{(k-1)}\leqslant X_{(n)}-X_{(n-1)}\}$$
$$=\int\int P\{X_{n+1},\dots,X_{n+s}\in(X_{(k-1)},X_{(k)}),X_{n+s+1},\dots,X_{n+m}\notin(X_{(k-1)},X_{(k)}),$$
$$X_{(k)}-X_{(k-1)}\leqslant X_{(1)},\dots,X_{(k)}-X_{(k-1)}\leqslant X_{(k-1)}-X_{(k-2)},$$
$$X_{(k)}-X_{(k-1)}\leqslant X_{(k+1)}-X_{(k)},\dots,X_{(k)}-X_{(k-1)}\leqslant X_{(n)}-X_{(n-1)}\mid$$
$$X_{(k)}-X_{(k-1)}=t_1,X_{(k-1)}=t_2\}$$
$$\times\mathrm{d}F_{X_{(k)}-X_{(k-1)}}(t_1)\mathrm{d}F_{X_{(k-1)}}(t_2)=\frac{n!}{(k-2)!(n-k)!}\lambda^2\int_0^\infty\int_0^\infty(\mathrm{e}^{-\lambda t_2}(1-\mathrm{e}^{-\lambda t_1}))^s$$
$$\times(1-(\mathrm{e}^{-\lambda t_2}(1-\mathrm{e}^{-\lambda t_1})))^{m-s}\mathrm{e}^{-\frac{n(n+1)}{2}\lambda t_1}(1-\mathrm{e}^{-\lambda t_2})^{k-2}(\mathrm{e}^{-\lambda t_2})^{n-k+2}\,\mathrm{d}t_1\,\mathrm{d}t_2$$
$$=\frac{n!(n(n+1)/2-1)!}{(n-k)!}\sum_{i=0}^{m-s}(-1)^i\binom{m-s}{i}\frac{(n-k+s+i+1)!}{(n+s+i)!}$$
$$\times\frac{(s+i)!}{(n(n+1)/2+s+i)!}.$$

Therefore,

$$P\{X_{n+1},\dots,X_{n+s}\in(X_{(v-1)},X_{(v)}),X_{n+s+1},\dots,X_{n+m}\notin(X_{(v-1)},X_{(v)})\}$$
$$=\frac{2}{n+1}\sum_{i=0}^{m-s}(-1)^i\binom{m-s}{i}\frac{n+s+i+1}{s+i+2}\binom{\frac{n(n+1)}{2}+s+i}{s+i}^{-1}.$$

The theorem is hence proved.

As a special cases of Theorem 3 we have the following corollary:

Corollary 1. *If* $F(x)=1-\exp(-\lambda x)$, $x\geqslant 0$, $\lambda>0$, *then*

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

and

$$P\{X_{n+1},X_{n+2},\dots,X_{n+m}\in(X_{(v-1)},X_{(v)})\}=\frac{(n(n+1)/2-1)!m!n(m+n+1)}{(n(n+1)/2+m)!(m+2)}.$$

Now, define the following random variables:

$$\xi_i=\begin{cases}1 & \text{if } X_{n+i}\in(X_{(v-1)},X_{(v)}),\\ 0 & \text{if } X_{n+i}\notin(X_{(v-1)},X_{(v)}),\end{cases}\qquad i=1,2,\dots,m,\quad X_{(0)}=0 \text{ and } S_m=\sum_{i=1}^{m}\xi_i.$$

Note that the random variables $\xi_1,\xi_2,\dots,\xi_n$ are dependent. It is evident that S_m is the number of $X_{n+1},X_{n+2},\dots,X_{n+m}$ falling into $(X_{(v-1)},X_{(v)})$. Remember that we denote by $A(m,n,s)$ the right-hand side in (2.1). We prove the following theorem.

Theorem 4. *Let* $F(x)=1-\exp(-\lambda x)$, $x\geqslant 0$, $\lambda>0$. *Then the distribution of random variable* S_m *is*

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

Proof. From the definition of S_m immediately follows:

$$P\{S_m=s\}=\sum_{i_1,\dots,i_m}P\{A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_s}\cap\overline{A_{i_{s+1}}}\cap\overline{A_{i_{s+2}}}\cap\cdots\cap\overline{A_{i_m}}\},\tag{2.3}$$

where $A_{i_j}=\{X_{n+i_j}\in(X_{(v-1)},X_{(v)})\}$, $j=1,2,\dots,m$ and $\overline{A_{i_j}}$ is the complement of event A_{i_j}.

The number of summands in (2.3) is equal to $\binom{m}{s}$ and each member of the sum has the same value $A(m,n,s)$.

The proof of the theorem is thus completed. □

3. Asymptotic distribution of S_m for continuous distributions

Let $X_1,X_2,\dots,X_n$ be a sample with a continuous distribution F and X_{n+1}, $X_{n+2},\dots,$ X_{n+m} be the next m observations obtained from the same population independent of $X_1,X_2,\dots,X_n$. Consider the spacings $X_{(1)}-X_{(0)}$, $X_{(2)}-X_{(1)}$, $X_{(3)}-X_{(2)},\dots,$ $X_{(n)}-X_{(n-1)}$ $(X_{(0)}=-\infty)$. Define a random variable v as in Section 2: $v=k$ iff $X_{(k)}-X_{(k-1)}\leqslant X_{(i)}-X_{(i-1)}$, $i=1,2,\dots,n$, i.e. v is the index of spacing having minimal length. Let S_m be defined as earlier, i.e. S_m be a number of observations $X_{n+1},X_{n+2},\dots,X_{n+m}$ falling into random interval $(X_{(v-1)},X_{(v)})$.

Theorem 5. *The asymptotic distribution of S_m/m for large m is*

$$\lim_{m\to\infty} \sup_{0\leqslant x\leqslant 1} \left| P\left\{ \frac{S_m}{m} \leqslant x \right\} - P\{F(X_{(v)}) - F(X_{(v-1)}) \leqslant x\} \right| = 0. \tag{3.1}$$

Proof. By definition

$$S_m = \sum_{i=1}^{m} \xi_i = \sum_{i=1}^{m} I_{(X_{(v-1)},X_{(v)})}(X_{n+i}), \tag{3.2}$$

where

$$I_B(x) = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{if } x \notin B. \end{cases}$$

Using representation (3.2) one has

$$\begin{aligned} P\left\{ \frac{S_m}{m} \leqslant x \right\} &= P\left\{ \frac{1}{m} \sum_{i=1}^{m} I_{(X_{(v-1)},X_{(v)})}(X_{n+i}) \leqslant x \right\} \\ &= P\left\{ \int_{-\infty}^{\infty} I_{(X_{(v-1)},X_{(v)})}(u)\, \mathrm{d}F_m^*(u) \leqslant x \right\}, \end{aligned} \tag{3.3}$$

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

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

$$G(F) = \int_{-\infty}^{\infty} I_{(X_{(k-1)},X_{(k)})}(u)\, \mathrm{d}F(u),$$

$G(F) = G(F)(\omega)$ is a random variable defined in a probability space $\{\Omega, \mathfrak{I}, P\}$. By using Glivenko–Cantelli theorem one can observe that $G^*(F_m^*) \to G^*(F)$ almost surely. From (3.3) it follows that

$$P\left\{ \frac{S_m}{m} \leqslant x \right\} = \sum_{k=1}^{n} P\left\{ \int_{-\infty}^{\infty} I_{(X_{(k-1)},X_{(k)})}(u)\, \mathrm{d}F_m^*(u) \leqslant x,\ v = k \right\}. \tag{3.4}$$

Repeating the same arguments for $\{v = k\}$ as in the proof of Theorem 3 and obtaining the limit as $m \to \infty$ in (3.4), one can see that (3.1) is an almost trivial consequence of Glivenko–Cantelli Theorem.

4. Distribution of statistics based on record values

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

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

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

It is known that the distribution of nth record value is

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

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) and Arnold et al. (1998) among others.

4.1. Distribution free properties

Let $X_{u(n)}$ be the nth record value of a sequence of i.i.d. r.v.'s. Suppose that $X_{u(n)+1}$, $X_{u(n)+2}, \ldots, X_{u(n)+m}$ be the m observations that succeed $X_{u(n)}$.

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

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

(cf., Bairamov, 1997).

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

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

Denote

$$v_m = \sum_{i=1}^{m} \xi_i,$$

v_m is the number of observations $X_{u(n)+1}, X_{u(n)+2}, \ldots, X_{u(n)+m}$ falling into interval $(X_{u(r)}, X_{u(s)})$. It is clear that the r.v.'s $\xi_1, \xi_2, \ldots, \xi_m$ are dependent.

Theorem 6. *For $r = k - 1$ and $s = k$ $(k = 2, \ldots, n)$,*

$$P\{v_m = j\} = \binom{m}{j} \sum_{i=0}^{m-j} (-1)^i \binom{m-j}{i} \frac{1}{(j+i+1)^k}, \quad j = 0, 1, \ldots, m.$$

Proof. From the definition of v_m it immediately follows that

$$P\{v_m = j\} = \sum_{i_1, \ldots, i_m} P\{A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j} \cap \overline{A_{i_{j+1}}} \cap \overline{A_{i_{j+2}}} \cap \cdots \cap \overline{A_{i_m}}\}, \tag{4.1}$$

where $A_{i_l} = \{X_{u(n)+i_l} \in (X_{u(k-1)}, X_{u(k)})\}$, $l = 1, 2, \ldots, m$ and $\overline{A_{i_l}}$ is the complement of event A_{i_l}.

$$
\begin{aligned}
&P\{A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s} \cap \overline{A_{i_{j+1}}} \cap \overline{A_{i_{j+2}}} \cap \cdots \cap \overline{A_{i_m}}\} \\
&\quad = P\{X_{u(n)+1} \in (X_{u(k-1)}, X_{u(k)}), \ldots, X_{u(n)+j} \in (X_{u(k-1)}, X_{u(k)}), \\
&\qquad X_{u(n)+j+1} \notin (X_{u(k-1)}, X_{u(k)}), \ldots, X_{u(n)+m} \notin (X_{u(k-1)}, X_{u(k)})\} \\
&\quad = \int\int P(X_{u(n)+1} \in (X_{u(k-1)}, X_{u(k)}), \ldots, X_{u(n)+j} \in (X_{u(k-1)}, X_{u(k)}), \\
&\qquad X_{u(n)+j+1} \notin (X_{u(k-1)}, X_{u(k)}), \ldots, X_{u(n)+m} \notin (X_{u(k-1)}, X_{u(k)}) \,|\, X_{u(k-1)} = t_1, X_{u(k)} = t_2) \\
&\qquad \times \mathrm{d}F_{k-1,k}(t_1, t_2) = \int_0^1 \int_{t_1}^1 (t_2 - t_1)^j (1 - (t_2 - t_1))^{m-j} \frac{1}{(k-2)!} \left[\ln \frac{1}{1 - t_1}\right]^{k-2} \\
&\qquad \times \frac{1}{1 - t_1} \,\mathrm{d}t_2 \,\mathrm{d}t_1 = \frac{1}{(k-2)!} \sum_{i=0}^{m-j} (-1)^i \binom{m-j}{i} \frac{1}{j+i+1} \\
&\qquad \times \left\{ \int_0^1 \left[\ln \frac{1}{1 - t_1}\right]^{k-2} (1 - t_1)^{j+i} \,\mathrm{d}t_1 \right\}.
\end{aligned} \tag{4.2}
$$

By changing variable $\ln 1/(1 - t_1) = z$ in (4.2) one has

$$
\begin{aligned}
&P\{A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s} \cap \overline{A_{i_{j+1}}} \cap \overline{A_{i_{j+2}}} \cap \cdots \cap \overline{A_{i_m}}\} \\
&\quad = \frac{1}{(k-2)!} \sum_{i=0}^{m-j} (-1)^i \binom{m-j}{i} \frac{1}{j+i+1} \left\{ \int_0^\infty z^{k-2} \mathrm{e}^{-z(j+i+1)} \,\mathrm{d}z \right\} \\
&\quad = \sum_{i=0}^{m-j} (-1)^i \binom{m-j}{i} \frac{1}{(j+i+1)^k}.
\end{aligned} \tag{4.3}
$$

The number of summands in (4.1) is equal to $\binom{m}{j}$ and all of them have the same probability (4.3), hence

$$
P\{v_m = j\} = \binom{m}{j} \sum_{i=0}^{m-j} (-1)^i \binom{m-j}{i} \frac{1}{(j+i+1)^k}, \quad j = 0, 1, \ldots, m.
$$

The assertion thus follows.

4.2. Exceedance statistics in a record threshold model

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

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

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 1. *It is true that*

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

Proof. One can write

$$S_m(r,s) = \sum_{i=1}^{m} \xi_i^*(r,s) = \sum_{i=1}^{m} I_{\{(X_{u(r)},X_{u(s)})\}}(X_i'), \tag{4.4}$$

where

$$I_{\{B\}}(x) = \begin{cases} 1 & \text{if } x\in B, \\ 0 & \text{if } x\notin B. \end{cases}$$

Using representation (4.4) one has

$$\begin{aligned} P\left\{\frac{S_m(r,s)}{m}\leqslant x\right\} &= P\left\{\frac{1}{m}\sum_{i=1}^{m} I_{\{(X_{u(r)},X_{u(s)})\}}(X_i')\leqslant x\right\} \\ &= P\left\{\int_{-\infty}^{\infty} I_{\{(X_{u(r)},X_{u(s)})\}}(u)\,\mathrm{d}F_m^*(u)\leqslant x\right\}, \end{aligned} \tag{4.5}$$

where $F_m^*(u)$ is the empirical distribution function of sample $X_1', X_2', \ldots, X_m'$. Denote

$$A^*(G) = \int_{-\infty}^{\infty} I_{\{(x,y)\}}(u)\,\mathrm{d}G(u), \tag{4.6}$$

$$A(G) = \int_{-\infty}^{\infty} I_{\{(X_{u(r)},X_{u(s)})\}}(u)\,\mathrm{d}G(u), \tag{4.7}$$

where $A(G)=A(G)(\omega)$ is a random variable defined in the probability space $\{\Omega,\mathfrak{J},P\}$. Using (4.7), one can write (4.5) as follows:

$$P\left\{\frac{S_m(r,s)}{m}\leqslant x\right\} = P\left\{\int_{-\infty}^{\infty} I_{\{(X_{u(r)},X_{u(s)})\}}(u)\,\mathrm{d}F_m^*(u)\leqslant x\right\} = P\left\{A(F_m^*)\leqslant x\right\}.$$

Functional (4.6) is continuous with respect to uniform metric. One can follow from the Glivenko–Cantelli Theorem ($P\{\omega\colon\ \sup_u |F_m^*(u)-G(u)| \to 0\}=1$) $A^*(F_m^*) \to A^*(G)$ almost surely. It is clear that

$$\begin{aligned} &P\left\{\omega\colon \lim_{m\to\infty} A(F_m^*) = A(G)\right\} \\ &= \int\!\!\int P\left\{\lim \int_{-\infty}^{\infty} I_{\{(X_{u(r)},X_{u(s)})\}}(u)\,\mathrm{d}F_m^*(u)\right. \\ &= \left.\int_{-\infty}^{\infty} I_{\{(X_{u(r)},X_{u(s)})\}}(u)\,\mathrm{d}G(u)\,\middle|\, X_{u(r)}=x, X_{u(s)}=y\right\}\,\mathrm{d}F_{r,s}(x,y) \\ &= \int\!\!\int_{x<y} P\left\{\lim \int_{-\infty}^{\infty} I_{\{(x,y)\}}(u)\,\mathrm{d}F_m^*(u) = \int_{-\infty}^{\infty} I_{\{(x,y)\}}(u)\,\mathrm{d}G(u)\right\}\,\mathrm{d}F_{r,s}(x,y) \\ &= \int\!\!\int_{x<y} P\{\lim A^*(F_m^*) = A^*(G)\}\,\mathrm{d}F_{r,s}(x,y) = 1, \end{aligned}$$

where $F_{r,s}(x,y)$ is the joint distribution function of $X_{u(r)}$ and $X_{u(s)}$. So $A(F_m^*) \to A(G)$ a.s. in $\{\Omega, \Im, P\}$. Thus $A(F_m^*) \to A(G)$ in distribution as $m \to \infty$. We have

$$P\{A(G) \leqslant x\} = P\left\{\int_{-\infty}^{\infty} I_{\{(X_{u(r)}, X_{u(s)})\}}(u)\, \mathrm{d}G(u) \leqslant x\right\}$$

$$= P\left\{\int_{X_{u(r)}}^{X_{u(s)}} \mathrm{d}G(u) \leqslant x\right\} = P\{G(X_{u(s)}) - G(X_{u(r)}) \leqslant x\}.$$

The lemma is hence proved.

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*

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

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

$$P\{U_{rs} \leqslant x\} = \frac{1}{(r-1)!(s-r-1)!} \int_0^x \int_0^{1-t_1} \left[\ln \frac{1}{1-t_2}\right]^{r-1} \frac{1}{1-t_2}$$

$$\times \left[\ln \frac{1-t_1}{1-t_1-t_2}\right]^{s-r-1} \mathrm{d}t_2\, \mathrm{d}t_1,$$

and p.d.f.

$$f_{U_{rs}}(x) = \frac{1}{(r-1)!(s-r-1)!} \int_0^{1-x} \left[\ln \frac{1}{1-t_2}\right]^{r-1}$$

$$\times \frac{1}{1-t_2}\left[\ln \frac{1-x}{1-x-t_2}\right]^{s-r-1} \mathrm{d}t_2, \quad 0 < x < 1.$$

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

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

$$\operatorname{Var} S_m(k-1,k) = m(1/2^k - 1/3^k) + m^2(1/3^k - 1/2^{2k}).$$

As a consequence of Lemma 1, we have the following:

Theorem 7. *Let $F = G$. For $r = k-1$ and $s = k$ $(k = 2, 3, \dots, n)$*

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

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

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

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

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

where

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

Acknowledgements

The authors thank the referee and an Editor for their valuable suggestions which improved the presentation of this paper. We also appreciate the valuable comments of Professor Yalçin Tuncer.

References

Ahsanullah, M., 1995. Record Statistics. Nova Science Publishers, Commack, NY.

Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1998. Records. Wiley, New York.

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., 1998. Asymptotic distributions of the statistics based on order statistics, record values and invariant confidence intervals. Proceedings of International Conference on Asymptotic Methods in Probability and Mathematical Statistics, St. Petersburg, June 24–28, 1998.

Bairamov, I.G., Petunin, Yu.I., 1991. Structure of invariant confidence intervals containing the main distributed mass. Theor. Probab. Appl. 35 (1), 15–26.

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 Edition. Klieger, Malabar, FL.

Gumbell, E.J., von Shelling, H., 1950. The distribution of number of exceedances. Ann. Math. Statist. 21, 247–262.

Johnson, N.L., Kotz, S., Kemp, A., 1992. Univariate Discrete Distributions. Wiley, New York.

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

Nagaraja, H.N., 1988. Record values and related statistics. A Review. Commun. Statist. Theory Methods 17, 2223–2238.

Nevzorov, V.B., 1988. Records. Theoret. Probab. Appl. 32, 201–228.

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

Sarkadi, K., 1957. On the distribution of the number of exceedances. Ann. Math. Statist. 28, 1021–1022.

Shewart, W.A., 1931. Economic Control of Quantity of Manufactured Product. Van Nostrand, Princeton, NJ.

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, pp. 417–423.

Wesolowsky, J., Ahsanullah, M., 1999. Distributional properties of exceedance statistics. Ann. Inst. Statist. Math. 50 (3), 543–565.

Wilks, S.S., 1941. On the determination of sample sizes for setting tolerance limits. Ann. Math. Statist. 12, 91–96.

Wilks, S.S., 1942. Statistical prediction with special references to the problem of tolerance limits. Ann. Math. Statist. 13, 400–409.