JOINT BEHAVIOUR OF PRECEDENCES AND EXCEEDANCES IN RANDOM THRESHOLD MODELS

ISMIHAN BAIRAMOV AND SERKAN ERYILMAZ*

Izmir University of Economics

Summary

This paper concerns the joint behaviour of precedence and exceedance statistics in random threshold models. Joint distributions of precedence and exceedance statistics, both exact and asymptotic, are obtained, and the results are illustrated for random thresholds based on order statistics and record values.

Key words: exceedance statistics; order statistics; precedence statistics; random threshold; record values.

1. Introduction

The problem of finding the distribution of precedence (exceedance) statistics, which denote the total number of observations preceding (exceeding) a threshold value, has always been the subject of much study. These kinds of statistics have been widely used in many areas involving statistical process control, reliability, and hydrological frequency analysis. If the threshold value is fixed, the problem is closely related to the classical binomial model. Random thresholds, however, make the problem more complicated and difficult. The use of a random threshold was first considered in the works of Gumbel & von Schelling (1950) and Epstein (1954).

Precedence and exceedance models based on random thresholds are generally set up using two random samples that are independent of each other. Random thresholds are determined according to one of the samples, and the behaviour of observations in the other sample with respect to these random thresholds is investigated. In this context, these statistics are potentially useful for testing whether the two random samples are from the same population: see, for example, Katzenbeisser (1989). We refer also to Balakrishnan & Ng (2006) for precedence type tests and applications.

The topic is closely related to tolerance limits (Robbins 1944), and to invariant confidence intervals containing the main distributed mass, as introduced by Bairamov & Petunin (1991). It is known that, under certain conditions, an invariant confidence interval for the class of all continuous distribution functions can be constructed only from the order statistics. As will be explained in the following sections, this result supports the use of order statistics as random thresholds.

Wesolowski & Ahsanullah (1998) investigated the distributional properties of various exceedance statistics. Bairamov & Eryilmaz (2000) obtained the distributions (both exact

and asymptotic) of exceedance-type statistics based on the minimal spacing of a sample and record values. Recent discussions on the topic can be found in Bairamov & Kotz (2001), Eryilmaz & Bairamov (2003), Stepanov (2004) and Bairamov & Eryilmaz (2006).

Here, we investigate the joint behaviour of precedence and exceedance statistics with respect to lower and upper random thresholds X_L and X_U. In Section 2, we present the joint distributions, both exact and asymptotic, of precedences and exceedances. Section 3 contains an illustration of distributions for random thresholds, based on order statistics and record values.

2. Joint distribution of precedence and exceedance statistics

Let $Y_1, Y_2, \ldots$ be independent, identically distributed (i.i.d.) random variables with a continuous cumulative distribution function (cdf) F. Let X_L and X_U be two random variables termed the lower and upper random thresholds, respectively. These two thresholds are generally dependent random variables, and may be viewed as a function of a random sample $X_1, X_2, \ldots$, which is independent of $Y_1, Y_2, \ldots$. Therefore, $X_L = \phi(X_1, X_2, \ldots)$ and $X_U = \psi(X_1, X_2, \ldots)$, where ϕ and ψ are two Borel functions satisfying $\phi(x_1, x_2, \ldots, x_m) \leq \psi(x_1, x_2, \ldots, x_m)$ for all $(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$.

Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables with continuous cdf G, and the joint cdf and probability density function (pdf) of X_L and X_U are denoted respectively by

$$H(x, y) = \Pr\{X_L \leq x, X_U \leq y\} \quad \text{and} \quad h(x, y) = \frac{\partial^2 H(x, y)}{\partial x \partial y}.$$

Define

$$S_n(X_L) = \#\{i \leq n : Y_i < X_L\} \quad \text{and} \quad S_n(X_U) = \#\{i \leq n : Y_i > X_U\}.$$

It is obvious that $S_n(X_L)$ ($S_n(X_U)$) denotes the number of Ys that precede (exceed) the random threshold X_L (X_U). The following result provides the joint probability mass function (pmf) of these random variables.

Theorem 1. *For $k + l \leq n$,*

$$\Pr[S_n(X_L) = k, S_n(X_U) = l]$$
$$= \binom{n}{k, l, n-k-l} E_H \left\{ F^k(X_L) \bar{F}^l(X_U) [F(X_U) - F(X_L)]^{n-k-l} \right\},$$

where $\binom{n}{k, l, n-k-l} = n!/\{k!l!(n-k-l)!\}$ and $\bar{F}(x) = 1 - F(x)$.

Proof. Let us introduce the following trivalue random variables:

$$\xi_i = \begin{cases} 0 & if \quad Y_i < X_L \\ 1 & if \quad Y_i \in (X_L, X_U), \quad i = 1, \ldots, n. \\ 2 & if \quad Y_i > X_U \end{cases}$$

Since the random variables $\xi_1, \xi_2, \ldots, \xi_n$ are exchangeable, one can write

$$\Pr[S_n(X_L) = k, S_n(X_U) = l]$$

$$= \binom{n}{k, l, n-k-l} \Pr\left[\sum_{i=1}^n I(\xi_i = 0) = k, \sum_{i=1}^n I(\xi_i = 2) = l, \right.$$

$$\left. \sum_{i=1}^n I(\xi_i = 1) = n-k-l \right],$$

where $I(A) = 1$ if A occurs and $I(A) = 0$ otherwise. The proof readily follows because $[\sum_{i=1}^n I(\xi_i = 0), \sum_{i=1}^n I(\xi_i = 2), \sum_{i=1}^n I(\xi_i = 1)]$ has a multinomial distribution of parameters $[n, p_1 = F(X_L), p_2 = \bar{F}(X_U), p_3 = F(X_U) - F(X_L)]$; that is, it is a mixture with the mixed distribution H. □

Now, consider the asymptotic behaviour of $(S_n(X_L)/n, S_n(X_U)/n)$ as n tends to infinity. Let us first establish the following limit result, which will be helpful.

Lemma 2. *For real values a, b, c and complex values α and β, if $a + b + c = 1$, then*

$$\lim_{n\to\infty} [a\exp(\alpha/n) + b\exp(\beta/n) + c]^n = \exp(a\alpha + b\beta).$$

Proof. One can write

$$n\log[a\exp(\alpha/n) + b\exp(\beta/n) + c]$$

$$= n\log\{a[1 + \alpha/n + o(1/n)] + b[1 + \beta/n + o(1/n)] + c\}$$

$$= n\log\left[1 + \frac{1}{n}(a\alpha + b\beta) + o(1/n)\right]$$

$$= n\left[\frac{a\alpha + b\beta}{n} + o(1/n)\right] \to a\alpha + b\beta \text{ as } n \to \infty. \qquad \square$$

We are now ready to construct the asymptotic distribution of $(S_n(X_L)/n, S_n(X_U)/n)$.

Theorem 3. *For $n \to \infty$,*

$$\left(\frac{S_n(X_L)}{n}, \frac{S_n(X_U)}{n}\right) \xrightarrow{d} (F(X_L), \bar{F}(X_U)).$$

Proof. The joint characteristic function of $S_n(X_L)/n$ and $S_n(X_U)/n$ is

$$\varphi_{S_n(X_L)/n, S_n(X_U)/n}(t_1, t_2) = \sum_{k+l\le n}\sum e^{it_1 k/n} e^{it_2 l/n} \Pr[S_n(X_L) = k, S_n(X_U) = l]$$

$$= E_H\left\{\sum_{k+l\le n}\sum \binom{n}{k, l, n-k-l} e^{it_1 k/n} e^{it_2 l/n} F^k(X_L)\bar{F}^l(X_U)\right.$$

$$\left. \times [F(X_U) - F(X_L)]^{n-k-l}\right\}$$

$$= E_H \left\{ \sum_{k+l \leq n} \sum \binom{n}{k, l, n-k-l} \left[F(X_L) e^{it_1/n} \right]^k \left[\bar{F}(X_U) e^{it_2/n} \right]^l \right.$$

$$\left. \times [F(X_U) - F(X_L)]^{n-k-l} \right\}$$

$$= E_H \left(\left\{ F(X_L) \exp(it_1/n) + \bar{F}(X_U) \exp(it_2/n) + [F(X_U) - F(X_L)] \right\}^n \right).$$

Using Lemma 2 with $a = F(X_L)$, $b = \bar{F}(X_U)$, $c = F(X_U) - F(X_L)$ and $\alpha = it_1$, $\beta = it_2$, one obtains (noting that taking the limit under the integral sign is allowed by the dominated convergence theorem)

$$\varphi_{S_n(X_L)/n, S_n(X_U)/n}(t_1, t_2) \underset{n \to \infty}{\to} E_H \left\{ \exp \left[it_1 F(X_L) + it_2 \bar{F}(X_U) \right] \right\},$$

and this completes the proof.   □

Note that the above theorem can be seen as a stochastic version of the law of large numbers, because it presents the convergence of a multinomial distribution of parameters $[n, p_1 = F(X_L), p_2 = \bar{F}(X_U), p_3 = F(X_U) - F(X_L)]$ in law to (p_1, p_2, p_3).

3. Precedences and exceedances based on order statistics and record values

In this section, we illustrate the findings of the previous section for the random thresholds defined with the help of order statistics and record values. Let $X_1, X_2, \ldots, X_m$ be i.i.d. random variables with cdf G, and denote by $X_{1:m} \leq X_{2:m} \leq \cdots \leq X_{m:m}$ the corresponding order statistics. Let

$$X_L = X_{r:m} \quad \text{and} \quad X_U = X_{s:m}, \quad 1 \leq r < s \leq m. \tag{1}$$

In this case, the corresponding precedence and exceedance statistics are defined respectively by

$$S_n(X_{r:m}) = \#\{i \leq n : Y_i < X_{r:m}\}, \quad S_n(X_{s:m}) = \#\{i \leq n : Y_i > X_{s:m}\},$$

where $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables with cdf F.

There is a strong motivation behind the choice of (1) in both theory and practice. It is well known that, under the hypothesis $H_0 : F = G$,

$$\Pr[Y_k \in (X_{r:m}, X_{s:m})] = \frac{s-r}{m+1}, \quad 1 \leq r < s \leq m, \tag{2}$$

and under H_0, $Y_1, Y_2, \ldots, Y_n$ can be seen as a continuation of $X_1, X_2, \ldots, X_m$ so that $Y_i \overset{d}{=} X_{m+i}, i = 1, \ldots, n$. In this case, the probability (2) can be rewritten as

$$\Pr[X_{m+1} \in (X_{r:m}, X_{s:m})] = \frac{s-r}{m+1},$$

which implies that $(X_{r:m}, X_{s:m})$ is a distribution free confidence interval containing the future observation in the class of all absolutely continuous distribution functions. It is also known that, if ϕ and ψ are two Borel functions satisfying $\phi(x_1, x_2, \ldots, x_m) \leq \psi(x_1, x_2, \ldots, x_m)$ for all $(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$, then the probability

$$\Pr\{X_{m+1} \in [\phi(X_1, X_2, \ldots, X_m), \psi(X_1, X_2, \ldots, X_m)]\}$$

is the same for all absolutely continuous distributions if and only if

$$\phi(X_1, X_2, \ldots, X_m) = X_{r:m} \quad \text{and} \quad \psi(X_1, X_2, \ldots, X_m) = X_{s:m}$$

for some $1 \leq r < s \leq m$ (see Bairamov & Petunin 1991).

These random thresholds are also of special importance in applications. For example, in statistical process control the random sample $X_1, X_2, \ldots, X_m$ is known to be a reference sample, and distribution free control charts proposed recently by Chakraborti, van der Laan & van de Wiel (2004) are based on order statistics of a reference sample with a lower control limit $X_{r:m}$ and upper control limit $X_{s:m}$.

Corollary 4. *Under the hypothesis $H_0 : F = G$, the joint pmf of $S_n(X_{r:m})$ and $S_n(X_{s:m})$ is*

$$\Pr[S_n(X_{r:m}) = k, S_n(X_{s:m}) = l] = \frac{\binom{n-k-l+s-r-1}{n-k-l}\binom{k+r-1}{k}\binom{m-s+l}{l}}{\binom{n+m}{m}}$$

$$k, l = 1, \ldots, n \; ; \; k+l \leq n.$$

Proof. Via Theorem (1),

$$\Pr[S_n(X_{r:m}) = k, S_n(X_{s:m}) = l]$$
$$= \binom{n}{k, l, n-k-l} \iint_{x<y} F^k(x)[1 - F(y)]^l$$
$$\times [F(y) - F(x)]^{n-k-l} g_{r,s}(x, y)\, dy dx, \tag{3}$$

where $g_{r,s}(x, y)$ denotes the joint pdf of $X_{r:m}, X_{s:m}$, and it is given by (see, for example, David & Nagaraja 2003)

$$g_{r,s}(x, y) = \binom{m}{r-1, s-r-1, m-s} G^{r-1}(x)\, [G(y) - G(x)]^{s-r-1}$$
$$\times [1 - G(y)]^{m-s} g(x)g(y), \quad x < y. \tag{4}$$

Using (4) in (3) and assuming $F = G$, one obtains

$$\Pr[S_n(X_{r:m}) = k, S_n(X_{s:m}) = l] = \binom{n}{k, l, n-k-l}\binom{m}{r-1, s-r-1, m-s}$$
$$\times \int_0^1 \int_u^1 u^{k+r-1}(v - u)^{n-k-l+s-r-1}(1 - v)^{m-s+l} dv du,$$

and simple manipulations yield the result. $\square$

Corollary 5. *For $n \to \infty$,*

$$\left(\frac{S_n(X_{r:m})}{n}, \frac{S_n(X_{s:m})}{n}\right) \xrightarrow{d} (F(X_{r:m}), \bar{F}(X_{s:m})), \tag{5}$$

and, under the hypothesis $H_0 : F = G$, the pdf of the vector $(F(X_{r:m}), \bar{F}(X_{s:m}))$ is

$$q(x, y) = \binom{m}{r-1, \, m-s, \, s-r-1} x^{r-1} y^{m-s} (1-x-y)^{s-r-1}$$

if $x + y < 1$, and $q(x, y) = 0$ if $x + y \geq 1$. The joint distribution given by the pdf $q(x, y)$ is actually the Dirichlet distribution with parameters r, $m - s + 1$, and $s - r$.

Proof. The result (5) directly follows from Theorem 3. For the second part, the probability integral transformation yields

$$\Pr \left[F(X_{r:m}) \leq x, \, \bar{F}(X_{s:m}) \leq y \right] = \Pr \left(U_{r:m} \leq x, \, U_{s:m} \geq 1 - y \right)$$

$$= \binom{m}{r-1, \, s-r-1, \, m-s} \int_0^x \int_{1-y}^1 u^{r-1} (v-u)^{s-r-1} (1-v)^{m-s} \, dv \, du, \qquad (6)$$

where $U_{r:m}$ and $U_{s:m}$ are the rth and sth order statistics of a random sample from the unit uniform distribution. Now, taking the second derivative of (6) with respect to y and x, the required result is obtained.

Distributional properties of precedences and exceedances can also be considered using record threshold models. These types of considerations are available in the literature. See for example Bairamov (1997), Wesolowski & Ahsanullah (1998) and Stepanov (2004).

Let $\{X_m\}_{m \geq 1}$ be a sequence of i.i.d. random variables with continuous cdf G. Define the record times of this sequence as $u(1) = 1$, $u(n) = \min \{ j : j > u(n-1), \, X_j > X_{u(n-1)} \}$, $n > 1$. $X_{u(1)}, X_{u(2)}, \ldots$ denote the corresponding record values. Now, let

$$X_L \equiv X_{u(r)} \quad \text{and} \quad X_U \equiv X_{u(s)}, \quad r < s.$$

The joint pdf of $X_{u(r)}$ and $X_{u(s)}$ is given by (see, for example, Ahsanullah 1995)

$$h(x, y) = \frac{R(x)^{r-1}}{(r-1)!} r(x) \frac{(R(y) - R(x))^{s-r-1}}{(s-r-1)!} g(y), \quad -\infty < x < y < \infty,$$

where

$$R(x) = -\ln(1 - G(x)) \quad \text{and} \quad r(x) = \frac{d}{dx} R(x) = \frac{g(x)}{1 - G(x)}. \qquad \square$$

In the following, we illustrate the joint distributions of precedence and exceedance statistics. We assume that $s = r + 1$ for simplicity.

Corollary 6. Under the hypothesis $H_0 : F = G$, the joint pmf of $S_n(X_{u(r)})$ and $S_n(X_{u(r+1)})$ is

$$\Pr[S_n(X_{u(r)}) = k, \, S_n(X_{u(r+1)}) = l]$$

$$= \frac{1}{n-k+1} \binom{n}{k} \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{1}{(j+n-k+1)^r}, \quad k + l \leq n.$$

Proof. Using Theorem 1, along with the joint pdf of $X_{u(r)}$ and $X_{u(r+1)}$,

$$\Pr\left[S_n(X_{u(r)}) = k, S_n(X_{u(r+1)}) = l\right]$$

$$= \binom{n}{k, l, n-k-l} \iint_{x<y} \left[F^k(x)(1 - F(y))^l (F(y) - F(x))^{n-k-l} \right.$$

$$\left. \times \frac{R(x)^{r-1}}{(r-1)!} \frac{1}{1 - F(x)} \, dF(x) dF(y) \right]$$

$$= \binom{n}{k, l, n-k-l} \int_0^1 \int_u^1 \left[u^k (1 - v)^l (v - u)^{n-k-l} \times \frac{(-\ln(1-u))^{r-1}}{(r-1)!} \frac{1}{1 - u} \right] dv du.$$

Using the transformation $1 - v = (1 - u)t$, the required result is obtained. □

Corollary 7. *For* $n \to \infty$,

$$\left(\frac{S_n(X_{u(r)})}{n}, \frac{S_n(X_{u(s)})}{n} \right) \xrightarrow{d} \left(F(X_{u(r)}), \bar{F}(X_{u(s)}) \right),$$

and, under the hypothesis $H_0 : F = G$, *the pdf of the vector* $(F(X_{u(r)}), \bar{F}(X_{u(s)}))$ *when* $s = r + 1$ *is*

$$z(x, y) = \frac{1}{(r-1)!} \frac{1}{1 - x} (-\ln(1-x))^{r-1}, \quad x + y < 1.$$

Proof. This immediately follows, using the joint distribution of $X_{u(r)}$ and $X_{u(s)}$. □

References

AHSANULLAH, M. (1995). *Record Statistics*. New York: Nova Science Publishers.

BAIRAMOV, I. (1997). Some distribution free properties of statistics based on record values and characterizations of the distributions through a record. *J. Appl. Statist. Sci.* **5**, 17–25.

BAIRAMOV, I. & ERYILMAZ, S. (2000). Distributional properties of statistics based on minimal spacing and record exceedance statistics. *J. Statist. Plann. Inference* **90**, 21–33.

BAIRAMOV, I. & ERYILMAZ, S. (2006). Spacings, exceedances and concomitants in progressive type II censoring scheme. *J. Statist. Plann. Inference* **136**, 527–536.

BAIRAMOV, I. & KOTZ, S. (2001). On distributions of exceedances associated with order statistics and record values for arbitrary distributions. *Statist. Papers* **42**, 171–185.

BAIRAMOV, I. & PETUNIN, Yu.I. (1991). Structure of invariant confidence intervals containing the main distributed mass. *Theor. Probab. Appl.* **35**, 15–26.

BALAKRISHNAN, N. & NG, H.K.T. (2006). *Precedence-type Tests and Applications*. Hoboken, NJ: John Wiley & Sons.

CHAKRABORTI, S., VAN DER LAAN, P. & VAN DE WIEL, M.A. (2004). A class of distribution free control charts. *J. Roy. Statist. Soc. Ser. C.* **53**, 443–462.

DAVID, H.A. & NAGARAJA, H.N. (2003). *Order Statistics*. New York: John Wiley & Sons.

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

ERYILMAZ, S. & BAIRAMOV, I. (2003). On a new sample rank of an order statistic and its concomitant. *Statist. Probab. Lett.* **63**, 123–131.

GUMBEL, E.J. & VON SCHELLING, H. (1950). The distribution of number of exceedances. *Ann. Math. Statist.* **21**, 247–262.

KATZENBEISSER, W. (1989). The exact power of two-sample location tests based on exceedance statistics against shift alternatives. *Statistics* **20**, 47–54.

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

STEPANOV, A. (2004). Random intervals based on records. *J. Statist. Plann. Inference* **118**, 103–113.

WESOLOWSKI, J. & AHSANULLAH, M. (1998). Distributional properties of exceedance statistics. *Ann. Inst. Statist. Math.* **50**, 543–565.