Journal of Computational and Applied Mathematics 236 (2012) 2259–2267

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

On asymptotic distributions of exceedance statistics

Nejla Ozkaya Turhan

Ankara University, Faculty of Science, Department of Statistics, 06100 Tandogan, Ankara, Turkey

ARTICLE INFO

Article history:
Received 15 March 2011
Received in revised form 5 September 2011

Keywords:
Exceedance statistics
Order statistics
Exact and asymptotic distribution
Random threshold

ABSTRACT

In this study, two independent samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ with respective distribution functions F and Q are considered. The joint asymptotic distributions of exceedance statistics defined as the number of Y observations falling into a random interval of order statistics constructed from the X sample is investigated. The results can be used in the context of a two-sample problem.

1. Introduction

Exceedance statistics have been widely used in survival analysis, reliability, construction of hypothesis testing especially for two-sample problems. Distributions of exceedance statistics are closely related to the tolerance and invariant confidence intervals containing the general distributed mass. Tolerance intervals were discussed in [1–3]. Invariant confidence intervals are discussed in [4]. A first discussion on exceedance statistics, their properties and applications can be found in [5,6]. Wesolowski and Ahsanullah [7] investigated the distributional properties of the various exceedance statistics. Matveychuk and Petunin [8] and Johnson and Kotz [9] used exceedance statistics for the construction of the two-sample hypothesis test. Katzenbeisser [10,11] proposed a test criteria by using exceedance statistics for testing whether the two random samples are from the same population or not. Bairamov and Petunin [4] introduced the notion of an invariant confidence interval containing the main distributed mass of a general population and showed that the only order statistics can be invariant confidence intervals for the class of all continuous distributions. Recently [12–14] studied exact and limiting distributions of exceedance statistics for both order statistics and record values when the underlying distribution is arbitrary.

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be samples obtained from populations with distribution functions F and Q, respectively, where $F, Q \in \Im_c$ and $\Im_c$ is the class of all continuous distribution functions. Let $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ and $Y_{1:m} \le Y_{2:m} \le \cdots \le Y_{m:m}$ be the order statistics constructed from $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$, respectively.

For any integer $m \ge 1$, define

$$S_m = \# \left\{ j \le m : Y_j \le X \right\} \tag{1}$$

which denotes the number of $Y_1, Y_2, \ldots, Y_m$ falling below the threshold X.

Denote by S_m^{rs} the number of Y observations falling into the random interval $(X_{r:n}, X_{s:n})$, i.e.

$$S_m^{rs} = \# \left\{ k \le m : Y_k \in (X_{r:n}, X_{s:n}) \right\} \quad 1 \le r < s \le n. \tag{2}$$

Denote

$$\triangledown_1 = (-\infty, X_{1:n}], \qquad \triangledown_2 = (X_{1:n}, X_{2:n}], \ldots, \qquad \triangledown_n = (X_{n-1:n}, X_{n:n}], \qquad \triangledown_{n+1} = (X_{n:n}, \infty).$$

E-mail address: nturhan@science.ankara.edu.tr.

0377-0427/$ – see front matter
doi:10.1016/j.cam.2011.11.014

and define binary random variables

$$\xi_i^k = \begin{cases} 1, & Y_k \in \triangledown_i \\ 0, & Y_k \notin \triangledown_i \end{cases}, k = 1, 2, \ldots, m; i = 1, 2, \ldots, n+1$$

$$S_i = \sum_{i=1}^{n+1} \xi_i^k, \quad k = 1, 2, \ldots, m.$$

It is obvious that S_i denotes the number of observations $Y_1, Y_2, \ldots, Y_m$ falling into interval $\triangledown_i$ $(i = 1, 2, \ldots, n+1)$.

Proposition 1.1. *Let the hypothesis H_0 be defined as $H_0 : F(x) = Q(x)$ and the composite alternative H_1 is $H_1 : F(x) \neq Q(x); F, Q \in \Im_c$. Assume that H_0 is true then, for $0 \leq i_k \leq m, k = 1, 2, \ldots, n+1, i_1 + i_2 + \cdots + i_{n+1} = m$ the joint probability mass function of random variables $\zeta_1, \zeta_2, \ldots, \zeta_{n+1}$ is*

$$P\{\zeta_1 = i_1, \zeta_2 = i_2, \ldots, \zeta_{n+1} = i_{n+1}\} = \frac{1}{\binom{n+m}{n}}.$$

Proof. If H_0 is true, then we have

$$\begin{aligned} P\{\zeta_1 = i_1, \zeta_2 = i_2, \ldots, \zeta_{n+1} = i_{n+1}\} &= \frac{n!m!}{i_1!i_2!\cdots i_{n+1}!} \int \cdots \int_{x_1 < x_2 < \cdots < x_n} F^{i_1}(x_1)\left[F(x_2) - F(x_1)\right]^{i_2} \cdots \\ &\quad \times \left[F(x_n) - F(x_{n-1})\right]^{i_n} dF(x_1)dF(x_2)\cdots dF(x_n) \\ &= \frac{n!m!}{i_1!i_2!\cdots i_{n+1}!} \frac{i_1!i_2!\cdots i_{n+1}!}{(i_1 + i_2 + \cdots + i_{n+1} + n)!} \\ &= \frac{1}{\binom{n+m}{n}}. \end{aligned} \tag{3}$$

It is not difficult to observe (3) can also be interpreted as follows: let the hypothesis H_0 be true and let $Z_1, Z_2, \ldots, Z_n$ be the ordered layout of the order statistics of both samples, then

$$P\{Z_1 < Z_2 < \cdots < Z_{n+m}\} = \frac{1}{\binom{n+m}{n}}.$$

For example

$$P\{X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n} \leq Y_{1:m} \leq Y_{2:m} \leq \cdots \leq Y_{m:m}\} = \frac{1}{\binom{n+m}{n}}.$$

(See [15, p. 442]). □

In this paper, the distribution and asymptotic behavior of the exceedance statistics S_1, S_2, S_3 defined as the number of observations of $Y's$ falling into $\triangle_1 = (-\infty, X_{1:n}]$, $\Delta_2 = (X_{1:n}, X_{n:n}]$, $\Delta_3 = (X_{n:n}, \infty)$ are studied respectively.

2. Joint distribution of exceedance statistics based on order statistics

Theorem 2.1. *Assume that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are samples from populations with d.f. F and Q, respectively where $F, Q \in \Im_c$. Let*

$$\triangle_1 = (-\infty, X_{1:n}], \qquad \Delta_2 = (X_{1:n}, X_{n:n}], \qquad \Delta_3 = (X_{n:n}, \infty)$$

and

$$\xi_i^k = \begin{cases} 1, & Y_k \in \Delta_i \\ 0, & Y_k \notin \Delta_i \end{cases} k = 1, 2, \ldots, m, i = 1, 2, 3.$$

Then

$$P\{S_1 = k, S_2 = l - k, S_3 = m - l\} = \frac{\binom{l-k+n-2}{n-2}}{\binom{m+n}{n}} \tag{4}$$

where $S_1 + S_2 + S_3 = m$, $S_1 \in [0, m]$, $S_2 \in [0, m]$, $S_3 \in [0, m]$.

Proof.

$$P\{S_1 = k, S_2 = l-k, S_3 = m-l\} = \sum_{i_1,i_2,\ldots,i_m} P\big\{Y_{i_1} \in \Delta_1, Y_{i_2} \in \Delta_1, \ldots, Y_{i_k} \in \Delta_1, Y_{i_{k+1}} \in \Delta_2, Y_{i_{k+2}} \in \Delta_2, \ldots, Y_{i_l} \in \Delta_2, Y_{i_{l+1}} \in \Delta_3, Y_{i_{l+2}} \in \Delta_3, \ldots, Y_{i_m} \in \Delta_3,\big\} \tag{5}$$

where the sum extends over all $m!$ permutations obtained from $1, 2, \ldots, m$. □

Denote by $F_{i,j}(x, y)$ the joint distribution function of order statistics $X_{i:n}$ and $X_{j:n}$. Applying the continuous total probability formula to (5), conditioning on $X_{1:n}$ and $X_{n:n}$ we have

$$\begin{aligned} &P\{S_1 = k, S_2 = l-k, S_3 = m-l\} \\ &= \frac{m!}{k!\,(l-k)!\,(m-l)!} \int\int_{-\infty<x_1<x_2<\infty} P\big\{Y_{i_1} \in (-\infty, X_{1:n}], \ldots, Y_{i_k} \in (-\infty, X_{1:n}], \\ &\quad Y_{i_{k+1}} \in (X_{1:n}, X_{n:n}], \ldots, Y_{i_l} \in (X_{1:n}, X_{n:n}], \\ &\quad Y_{i_{l+1}} \in (X_{n:n}, \infty), \ldots, Y_m \in (X_{n:n}, \infty) \mid X_{1:n} = x_1, X_{n:n} = x_2\big\}\, dF_{1,n}(x_1, x_2) \\ &= \frac{m!}{k!\,(l-k)!\,(m-l)!} \frac{n!}{(n-2)!} \int_{-\infty}^{\infty}\int_{-\infty}^{x_2} [F(x_1)]^k [F(x_2) - F(x_1)]^{l-k+n-2} [1 - F(x_2)]^{m-l} f(x_2) f(x_1) dx_1 dx_2 \\ &= \frac{\binom{l-k+n-2}{n-2}}{\binom{m+n}{n}}. \end{aligned}$$

Thus, the proof is completed.

3. Asymptotic distributions of exceedance statistics

Lemma 3.1.

$$\lim_{n\to\infty} \sup_{0\le x\le 1} \left| P\left\{\frac{S_m^{rs}}{m} \le x\right\} - P\{Q(X_{s:n}) - Q(X_{r:n}) \le x\} \right| = 0$$

for r-th and s-th order statistics ($1 \le r < s \le n$).

Following results are given for the $F = Q$ case. (see [16,17]).

Corollary 3.1. *Let $H_0 : F(x) = Q(x)$. Then*

$$\lim_{n\to\infty} \sup_{0\le x\le 1} \left| P\left\{\frac{S_m^{rs}}{m} \le x\right\} - P\{F(X_{s:n}) - F(X_{r:n}) \le x\} \right| = 0$$

Bairamov [16]. Let $W_{rs} = F(X_{s:n}) - F(X_{r:n})$, then the probability density function of W_{rs} is

$$f_{W_{rs}}(x) = \frac{1}{B(s-r, n-s+r+1)} x^{s-r-1} (1-x)^{n-s+r}, \quad 0 \le x \le 1$$

David [18].

Theorem 3.1. *The asymptotic distribution of $(S_1/m, S_2/m, S_3/m)$ is*

$$\begin{aligned} \lim_{m\to\infty} P\left\{\frac{S_1}{m} \le x_1, \frac{S_2}{m} \le x_2, \frac{S_3}{m} \le x_3\right\} &= \lim_{m\to\infty} P\{S_1 \le [mx_1], S_2 \le [mx_2], S_3 \le [mx_3]\} \\ &= (x_2)^n - (x_2 - x_1)^n - n x_1 (1 - x_3)^{n-1}. \end{aligned} \tag{6}$$

Proof. By substituting $S_3 = m - S_1 - S_2$ in (6) we reach

$$\begin{aligned} P\{S_1 \le mx_1, S_2 \le mx_2, S_3 \le mx_3\} &= P\{S_1 \le [mx_1], S_2 \le [mx_2], m - [mx_3] \le S_1 + S_2\} \\ &= \sum_{k=0}^{[mx_1]} \sum_{l=m-[mx_3]+k}^{[mx_2]} \frac{\binom{l-k+n-2}{n-2}}{\binom{m+n}{n}} \end{aligned}$$

where $[x]$ denotes the integer part of x. Let $t = l - k$

$$P\{S_1 \le mx_1, S_2 \le mx_2, S_3 \le mx_3\} = \frac{1}{\binom{m+n}{n}} \sum_{k=0}^{[mx_1]} \sum_{t=m-[mx_3]}^{[mx_2]-k} \binom{t+n-2}{n-2}$$

$$= \frac{1}{\binom{m+n}{n}} \sum_{k=0}^{[mx_1]} \left(\sum_{t=0}^{[mx_2]-k} \binom{t+n-2}{n-2} - \sum_{t=0}^{m-[mx_3]-1} \binom{t+n-2}{n-2} \right)$$

$$= \frac{1}{\binom{m+n}{n}} \left(\sum_{k=0}^{[mx_1]} \binom{[mx_2]-k+n-2+1}{[mx_2]-k} - \sum_{k=0}^{[mx_1]} \binom{m-[mx_3]-1+n-2+1}{m-[mx_3]-1} \right)$$

$$= \frac{1}{C_{n+m}^{n}} \left(\sum_{k=0}^{[mx_1]} \frac{([mx_2]-k+n-1)!}{([mx_2]-k)!\,(n-1)!} - \sum_{k=0}^{mx_1} \frac{(m-[mx_3]+n-2)!}{(m-[mx_3]-1)!\,(n-1)!} \right)$$

$$= \frac{1}{C_{n+m}^{n}} \left[\underbrace{\frac{1}{(n-1)!} \sum_{k=0}^{[mx_1]} \frac{([mx_2]-k+n-1)!}{([mx_2]-k)!}}_{I_1} - \underbrace{\frac{1}{(n-1)!} \sum_{k=0}^{[mx_1]} \frac{(m-[mx_3]-k+n-2)!}{(m-[mx_3]-k-1)!}}_{I_2} \right].$$

On the other hand, I_1 is simplified using the equations

$$\sum_{k=n_1}^{n_2} a_k = \sum_{k=n-n_2}^{n-n_1} a_{n-k} \tag{7}$$

$$\sum_{t=1}^{n} t(t+1)\cdots(t+m) = \frac{1}{m+2} n\,(n+1)\cdots(n+m+1). \tag{8}$$

$$\begin{aligned}
I_1 &= \frac{1}{(n-1)!} \sum_{k=0}^{[mx_1]} \frac{([mx_2]-k+n-1)!}{([mx_2]-k)!} \\
&= \frac{1}{(n-1)!} \sum_{k=0}^{[mx_1]} ([mx_2]-k+1)\,([mx_2]-k+2)\cdots([mx_2]-k+n-1) \\
&= \frac{1}{(n-1)!} \sum_{k=[mx_2-mx_1]}^{[mx_2]} (k+1)\,(k+2)\cdots(k+n-1) \\
&= \frac{1}{(n-1)!} \left[\sum_{k=1}^{[mx_2+1]} k\,(k+1)\,(k+2)\cdots(k+n-2) - \sum_{k=1}^{[mx_2-mx_1]} k\,(k+1)\,(k+2)\cdots(k+n-2) \right] \\
&= \frac{1}{n!} \left[([mx_2]+1)\,([mx_2]+2)\cdots([mx_2]+n) - ([mx_2-mx_1]) \right. \\
&\quad \left. \times\ ([mx_2]-[mx_1]+1)\cdots([mx_2]-[mx_1]+n-1)\right]
\end{aligned} \tag{9}$$

$$\begin{aligned}
I_2 &= \frac{1}{(n-1)!} \sum_{k=0}^{[mx_1]} \frac{(m-[mx_3]+n-2)!}{(m-[mx_3]-1)!} \\
&= \frac{1}{(n-1)!} \sum_{k=0}^{[mx_1]} (m-[mx_3])\,(m-[mx_3]+1)\cdots(m-[mx_3]+n-2) \\
&= \frac{1}{(n-1)!} ([mx_1]+1)\,(m-[mx_3])\,(m-[mx_3]+1)\cdots(m-[mx_3]+n-2).
\end{aligned}$$

Under the simplifications of I_1 and I_2, the joint distribution function of S_1/m, S_2/m and S_3/m is obtained as,

$$\begin{aligned}
P\{S_1 \le mx_1, S_2 \le mx_2, S_3 \le mx_3\} &= I_1 - I_2 \\
&= \frac{1}{\binom{m+n}{n}} \frac{1}{n!} \left([mx_2]+1\right)\left([mx_2]+2\right)\cdots\left([mx_2]+n\right)
\end{aligned}$$

$$- \frac{1}{n!}\left([mx_2 - mx_1]\right)\left([mx_2 - mx_1] + 1\right)\cdots\left([mx_2 - mx_1] + n - 1\right)$$
$$- \frac{1}{(n-1)!}\left(\left([mx_1] + 1\right)\left(m - [mx_3]\right)\right.$$
$$\left.\times \left(m - [mx_3] + 1\right)\cdots\left(m - [mx_3] + n - 2\right)\right) \tag{10}$$

where $[mx_1] \le [mx_2]$, $m - [mx_3] \le [mx_2]$, $m - [mx_3] + [mx_1] \le [mx_2]$.

By applying limits to $P\{S_1 \le mx_1, S_2 \le mx_2, S_3 \le mx_3\}$ in (10), the result is

$$\lim_{m\to\infty} P\{S_1 \le mx_1, S_2 \le mx_2, S_3 \le mx_3\} = (x_2)^n - (x_2 - x_1)^n - nx_1(1 - x_3)^{n-1} \tag{11}$$

where $x_1, x_2, x_3 \in [0, 1]$. As a special case of (11) for $n = 2$

$$\lim_{m\to\infty} P\{S_1 \le mx_1, S_2 \le mx_2, S_3 \le mx_3\} = (x_2)^2 - (x_2 - x_1)^2 - 2x_1(1 - x_3)^1$$

where $x_1, x_2, x_3 \in [0, 1]$. When S_1, S_2 and $S_3 \in [0, m]$ the probability is verified in the sense that

$$\begin{aligned} P\{S_1 \le m, S_2 \le m, S_3 \le m\} &= \frac{1}{\binom{m+n}{n}} \sum_{k=0}^{m} \sum_{t=0}^{m-k} \binom{t+n-2}{n-2} \\ &= \frac{n!m!}{(m+n)!} \frac{1}{n!} (m+1)(m+2)\cdots(m+n) \\ &= 1. \quad \square \end{aligned}$$

3.1. Asymptotic distribution of exceedance statistics when the random threshold is $X_{n:n}$

Assume that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be samples from populations with d.f. F and Q respectively where $F, Q \in \Im_c$. Let $X_{n:n}$ is a random threshold and $\Delta_1 = (-\infty, X_{n:n}]$, $\Delta_2 = (X_{n:n}, \infty)$ are intervals. And let us define the variables

$$\xi_i^k = \begin{cases} 1, & Y_k \in \Delta_i \\ 0, & Y_k \notin \Delta_i \end{cases} k = 1, 2, \ldots, m, i = 1, 2$$

and

$$S_i = \sum_{k=1}^{m} \xi_i^k.$$

Let S_1 and S_2 be exceedances, obtained from the sample $Y_1, Y_2, \ldots, Y_m$, denoting the numbers of observations below and above the random threshold $X_{n:n}$ respectively where $S_1 + S_2 = m$, $S_1 \in [0, m]$ and $S_2 \in [0, m]$.

Theorem 3.2. *Let $S_1 \in [0, m]$ and $S_2 \in [0, m]$ where $S_1 + S_2 = m$. Then the joint asymptotic distribution of $\frac{S_1}{m}$ and $\frac{S_2}{m}$ random variables is*

$$\lim_{m\longrightarrow\infty} P\left\{\frac{S_1}{m} \le x_1, \frac{S_2}{m} \le x_2\right\} = x_1^n - (1 - x_2)^n, \quad x_1 + x_2 > 1$$
$$\lim_{m\longrightarrow\infty} P\left\{\frac{S_1}{m} \le x_1\right\} = x_1^n, \quad x_1 \in [0, 1], x_2 > 1$$
$$\lim_{m\longrightarrow\infty} P\left\{\frac{S_2}{m} \le x_2\right\} = x_2^n, \quad x_2 \in [0, 1], x_1 > 1.$$

Proof. The joint probability function of S_1 and S_2 is

$$\begin{aligned} P\{S_1 = k, S_2 = m - k\} &= \sum_{i_1, i_2, \ldots, i_m} P\left\{Y_{i_1} \in \Delta_1, Y_{i_2} \in \Delta_1, \ldots, Y_{i_k} \in \Delta_1, Y_{i_{k+1}} \in \Delta_2, Y_{i_{k+2}} \in \Delta_2, \ldots, Y_{i_m} \in \Delta_2\right\} \\ &= \frac{m!}{k!(m-k)!} \int P\left\{Y_{i_1} \in (-\infty, x], Y_{i_2} \in (-\infty, x], \ldots,\right. \\ &\quad \left. Y_{i_k} \in (-\infty, x], Y_{i_{k+1}} \in (x, \infty), Y_{i_{k+2}} \in (x, \infty), \ldots, Y_{i_m} \in (x, \infty)\right\} n[F(x)]^{n-1} f(x)\,dx \\ &= \frac{m!}{k!(m-k)!} n \int_0^1 t^{n+k-1} [1 - t]^{m-k}\,dt \end{aligned}$$

$$
\begin{aligned}
&= \frac{n\,(n+k-1)!\,(m-k)!}{(m+n)!}\,\frac{m!}{k!\,(m-k)!} \\
&= \frac{nm!}{(m+n)!}\,(k+1)\,(k+2)\cdots(n+k-1)\,. \qquad (12)
\end{aligned}
$$

The marginal distribution function of $\frac{S_1}{m}$ is

$$
\begin{aligned}
P\left\{\frac{S_1}{m} \le x_1\right\} &= \sum_{k=0}^{[mx_1]} \frac{nm!}{(n+m)!}(k+1)(k+2)\cdots(n+k-1) \\
&= \sum_{k=1}^{[mx_1+1]} \frac{nm!}{(n+m)!}k(k+1)(k+2)\cdots(n+k-2) \\
&= \frac{nm!}{(n+m)!}\left[\frac{1}{n-2+2}\,[mx_1+1]\,[mx_1+2]\cdots[mx_1+n]\right] \\
&= \frac{1}{(m+1)\,(m+2)\cdots(m+n)}\,[mx_1+1]\,[mx_1+2]\cdots[mx_1+n]\,. \qquad (13)
\end{aligned}
$$

The marginal distribution function $\frac{S_2}{m}$ is obtained by using the same procedure in Eq. (13)

$$
\begin{aligned}
P\left\{\frac{S_2}{m} \le x_2\right\} &= \sum_{\substack{k=0\\k+i=m}}^{[mx_2]} \sum_{i=0}^{m} P\{S_1 = k, S_2 = i\} \\
&= \sum_{k=0}^{[mx_2]} \frac{nm!}{(n+m)!}\,\frac{(n+k-1)!}{k!} \\
&= \frac{1}{(m+1)\,(m+2)\cdots(m+n)}\,[mx_2+1]\,[mx_2+2]\cdots[mx_2+n]\,. \qquad (14)
\end{aligned}
$$

The joint distribution of the random variables $\frac{S_1}{m}$ and $\frac{S_2}{m}$ is

$$
\begin{aligned}
P\left\{\frac{S_1}{m} \le x_1, \frac{S_2}{m} \le x_2\right\} &= P\{S_1 \le mx_1, m - S_1 \le mx_2\} = P\{m - mx_2 \le S_1 \le mx_1\} \\
&= \sum_{k=[m-mx_2]}^{[mx_1]} P\{S_1 = k\} = \sum_{\substack{k=[m-mx_2]\\k+i=m}}^{[mx_1]} \sum_{i=0}^{m} P\{S_1 = k, S_2 = i\} \\
&= \sum_{k=[m-mx_2]}^{[mx_1]} \frac{nm!}{(n+m)!}\,\frac{(n+k-1)!}{k!} \\
&= \frac{nm!}{(n+m)!} \sum_{k=[m-mx_2]}^{[mx_1]} (k+1)(k+2)\cdots(k+n-1) \\
&= \frac{nm!}{(n+m)!}\left[\frac{1}{n-2+2}\,[mx_1+1]\,[mx_1+2]\,\cdots[mx_1+1+n-2+1]\right. \\
&\qquad \left. - \frac{1}{n-2+2}\,[m-mx_2]\,[m-mx_2+1]\,\cdots[m-mx_2+n-2+1]\right]. \qquad (15)
\end{aligned}
$$

The equations in (13)–(15) are simplified by (8). Thus, we derived the asymptotic distribution of $\frac{S_1}{m}$ and $\frac{S_2}{m}$ by using (13)–(15) when $m \to \infty$.

$$
\begin{aligned}
&\lim_{m\longrightarrow\infty} P\left\{\frac{S_1}{m} \le x_1, \frac{S_2}{m} \le x_2\right\} = x_1^n - (1-x_2)^n\,, \quad x_1 + x_2 > 1 \\
&\lim_{m\longrightarrow\infty} P\left\{\frac{S_1}{m} \le x_1\right\} = x_1^n, \quad x_1 \in [0, 1]\,, x_2 > 1 \\
&\lim_{m\longrightarrow\infty} P\left\{\frac{S_2}{m} \le x_2\right\} = x_2^n, \quad x_2 \in [0, 1]\,, x_1 > 1
\end{aligned}
$$

Thus, the proof is completed. □

3.2. Asymptotic distribution of exceedance statistics when the random threshold is $X_{1:n}$

Assume that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be samples from populations with d.f. F and Q respectively where $F, Q \in \Im_c$. Let $X_{1:n}$ is a random threshold and $\Delta_1 = (-\infty, X_{1:n}]$, $\Delta_2 = (X_{1:n}, \infty)$ are intervals. And let us define the variables

$$\xi_i^k = \begin{cases} 1, & Y_k \in \Delta_i \\ 0, & Y_k \notin \Delta_i \end{cases} k = 1, 2, \ldots, m, i = 1, 2$$

and

$$S_i = \sum_{k=1}^{m} \xi_i^k.$$

Let $S_1$and S_2 be exceedances, obtained from the sample $Y_1, Y_2, \ldots, Y_m$, denoting the numbers of observations below and above the random threshold $X_{1:n}$ respectively where $S_1 + S_2 = m$, $S_1 \in [0, m]$ and $S_2 \in [0, m]$.

Theorem 3.3. *Let $S_1 \in [0, m]$ and $S_2 \in [0, m]$, where $S_1 + S_2 = m$. Then the joint asymptotic distribution of $\frac{S_1}{m}$ and $\frac{S_2}{m}$ random variables is*

$$\lim_{m \longrightarrow \infty} P\left\{\frac{S_1}{m} \le x_1, \frac{S_2}{m} \le x_2\right\} = x_2^n - (1 - x_1)^n, \quad x_1, x_2 \in [0, 1], x_1 + x_2 > 1$$

$$\lim_{m \longrightarrow \infty} P\left\{\frac{S_1}{m} \le x_1\right\} = 1 - (1 - x_1)^n, \quad x_1 \in [0, 1], x_2 > 1$$

$$\lim_{m \longrightarrow \infty} P\left\{\frac{S_2}{m} \le x_2\right\} = 1 - (1 - x_2)^n, \quad x_2 \in [0, 1], x_1 > 1.$$

Proof. The joint probability function of S_1 and S_2 is

$$\begin{aligned} P\{S_1 = k, S_2 = m - k\} &= \sum_{i_1, i_2, \ldots, i_m} P\left\{Y_{i_1} \in \Delta_1, Y_{i_2} \in \Delta_1, \ldots, Y_{i_k} \in \Delta_1, Y_{i_{k+1}} \in \Delta_2, Y_{i_{k+2}} \in \Delta_2, \ldots, Y_{i_m} \in \Delta_2\right\} \\ &= \frac{m!}{k!\,(m-k)!} n \int_{-\infty}^{\infty} [F(x)]^k [1 - F(x)]^{m+n-k-1} \, dF(x) \\ &= \frac{nm!}{(m+n)!} (m - k + 1)(m - k + 2) \cdots (m + n - k - 1). \end{aligned} \tag{16}$$

The marginal distribution function of $\frac{S_1}{m}$ is

$$\begin{aligned} P\left\{\frac{S_1}{m} \le x_1\right\} &= \sum_{k=0}^{[mx_1]} \frac{nm!}{(n+m)!} (m-k+1)(m-k+2) \cdots (m+n-k-1) \\ &= \frac{nm!}{(n+m)!} \sum_{t=[m-mx_1]}^{m} (t+1)(t+2) \cdots (n+t-1) \\ &= 1 - \frac{[m - mx_1][m - mx_1 + 1] \cdots [m - mx_1 + n - 2 + 1]}{(m+1)(m+2) \cdots (m+n)}. \end{aligned} \tag{17}$$

The marginal distribution function $\frac{S_2}{m}$ is obtained by using the same procedure in Eq. (17)

$$\begin{aligned} P\left\{\frac{S_2}{m} \le x_2\right\} &= \sum_{k=0}^{[mx_2]} P\{S_2 = k\} \\ &= \sum_{k=0}^{[mx_2]} \frac{nm!}{(n+m)!} \frac{(n+m-k-1)!}{(m-k)!} \\ &= 1 - \frac{[m - mx_2][m - mx_2 + 1] \cdots [m - mx_2 + n - 2 + 1]}{(m+1)(m+2) \cdots (m+n)}. \end{aligned} \tag{18}$$

The joint distribution function of the random variables $\frac{S_1}{m}$ and $\frac{S_2}{m}$ is

$$\begin{aligned} P\left\{\frac{S_1}{m} \le x_1, \frac{S_2}{m} \le x_2\right\} &= P\{S_1 \le mx_1, m - S_1 \le mx_2\} \\ &= P\{m - mx_2 \le S_1 \le mx_1\} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=[m-mx_2]}^{[mx_1]} P\{S_1 = k\} \\
&= \sum_{\substack{k=[m-mx_2] \\ k+i=m}}^{[mx_1]} \sum_{i=0}^{m} P\{S_1 = k, S_2 = i\} \\
&= \sum_{k=[m-mx_2]}^{[mx_1]} \frac{nm!}{(n+m)!} \frac{(n+m-k-1)!}{(m-k)!} \\
&= \frac{nm!}{(n+m)!} \sum_{t=[m-mx_1]}^{[m-m+mx_2]} (t+1)(t+2)\cdots(t+n-1) \\
&= \frac{1}{(m+1)(m+2)\cdots(m+n)} [[mx_2+1][mx_2+2] \\
&\quad \cdots [mx_2+1+n-1] - [m-mx_1][m-mx_1+1]\cdots[m-mx_1+n-1]]. \qquad (19)
\end{aligned}
$$

The equations in (17)–(19) are simplified by (7) and (8). Thus, we derived asymptotic distribution of $\frac{S_1}{m}$ and $\frac{S_2}{m}$ by using (17)–(19) when $m \to \infty$.

$$\lim_{m\longrightarrow\infty} P\left\{\frac{S_1}{m} \le x_1, \frac{S_2}{m} \le x_2\right\} = x_2^n - (1-x_1)^n, \quad x_1, x_2 \in [0,1], x_1 + x_2 > 1$$

$$\lim_{m\longrightarrow\infty} P\left\{\frac{S_1}{m} \le x_1\right\} = 1 - (1-x_1)^n, \quad x_1 \in [0,1], x_2 > 1$$

$$\lim_{m\longrightarrow\infty} P\left\{\frac{S_2}{m} \le x_2\right\} = 1 - (1-x_2)^n, \quad x_2 \in [0,1], x_1 > 1$$

Thus, the proof is completed. □

4. Numerical results

In the following we provide numerical values for the exact distribution for different values of n and m.

$n = 3, m = 4$					
$l \backslash k$	0	1	2	3	4
0	0.02857				
1	0.05714	0.02857			
2	0.08571	0.05714	0.02857		
3	0.11429	0.08571	0.05714	0.02857	
4	0.14286	0.11429	0.08571	0.05714	0.02857

$n = 5, m = 4$					
$l \backslash k$	0	1	2	3	4
0	0.00794				
1	0.03175	0.00794			
2	0.07937	0.03175	0.00794		
3	0.15873	0.07937	0.03175	0.00794	
4	0.27778	0.15873	0.07937	0.03175	0.00794

$n = 4, m = 5$						
$l \backslash k$	0	1	2	3	4	5
0	0.00794					
1	0.02381	0.00794				
2	0.04762	0.02381	0.00794			
3	0.07937	0.04762	0.02381	0.00794		
4	0.11905	0.07937	0.04762	0.02381	0.00794	
5	0.16667	0.11905	0.07937	0.04762	0.02381	0.00794

$n = 4, m = 7$								
$l \backslash k$	0	1	2	3	4	5	6	7
0	0.00303							
1	0.00909	0.00303						
2	0.01818	0.00909	0.00303					
3	0.03030	0.01818	0.00909	0.00303				
4	0.04545	0.03030	0.01818	0.00909	0.00303			
5	0.06364	0.04545	0.03030	0.01818	0.00909	0.00303		
6	0.08485	0.06364	0.04545	0.03030	0.01818	0.00909	0.00303	
7	0.10909	0.08485	0.06364	0.04545	0.03030	0.01818	0.00909	0.00303

These results could be advised for use in two-sample hypothesis tests in many areas as described in the Introduction section.

Acknowledgments

The author would like to thank two anonymous reviewers for their invaluable comments.

References

[1] S.S. Wilks, On the determination of sample sizes for setting tolerance limits, The Annals of Mathematical Statistics 12 (1941) 91–96.
[2] S.S. Wilks, Statistical prediction with special references to the problem of the tolerance limits, Annals of mathematical Statistics 13 (1942) 400–440.
[3] H. Robbins, On distribution free tolerance limits in random sampling, The Annals of Mathematical Statistics 15 (1944) 214–216.
[4] I.G. Bairamov, YU.I. Petunin, Structure of invariant confidence intervals containing the main distributed mass, Theory of Probability and its Applications 35 (1) (1991) 15–26.
[5] E.J. Gumbel, H. von Shelling, The distribution of the number of exceedances, The Annals of Mathematical Statistics 21 (1950) 247–262.
[6] B. Epstein, Tables for the distribution of the number of exceedances, The Annals of Mathematical Statistics 25 (1954) 762–768.
[7] J. Wesolowski, M. Ahsanullah, Distributional properties of exceedance statistics, The Annals of Mathematical Statistics 50 (3) (1998) 543–565.
[8] S.A. Matveychuk, I. Yu Petunin, A generalization of the Bernoulli model arising in order statistics, I/II Ukrainian Mathematical Journal 42 (518–528/43) (1990/1991) 779–785.
[9] N.L. Johnson, S. Kotz, Further comments on Matveychuk and Petunin's generalized Bernoulli model, and nonparametric tests of homogeneity, Journal of Statistical Planning and Inference 41 (1994) 61–72.
[10] W. Katzenbeisser, The distribution of two sample location exceedance test statistics under Lehmann alternatives, Statistics Hefte (now Papers) 26 (1985) 131–138.
[11] W. Katzenbeisser, The exact power of two-sample location tests based on exceedances against shift alternatives, Mathematische Operationsforschung und Statistik 20 (1989) 47–54.
[12] I.G. Bairamov, S. Kotz, On distributions of exceedances associated with order statistics and record values for arbitrary distributions, Statistical Papers 42 (2) (2001) 171–185.
[13] I. Bairamov, M.K. Khan, On exceedances of record and order statistics, Proceedings of the American Mathematical Society 135 (6) (2007) 1935–1945.
[14] I. Bairamov, S. Eryilmaz, Joint behavior of preceedances and exceedances in random threshold models, Australian and New Zeland Journal of Statistics (2008) 209–216.
[15] S.S. Wilks, Mathematical Statistics, John Wiley & Sons, Inc, New York–London, 1962.
[16] I.G. Bairamov, Some distribution free properties of statistics based on record values and characterizations of the distributions through a record, Journal of Applied Statistical Science 5 (1) (1997) 17–25.
[17] I.G. Bairamov, O.L. Gebizlioğlu, 1998, Asymptotic distributions of statistics based on order statistics and record values, International conference, Asymptotical Methods in Probability and statistics, St. Petersburg, pp. 20–23.
[18] H.A. David, Order Statistics, Wiley, New York, 1981.