TWMS J. Pure Appl. Math. V.11, N.2, 2020, pp.189-203

BIVARIATE GENERAL RANDOM THRESHOLD MODELS AND EXCEEDANCE STATISTICS

AYSEGUL EREM[1], ISMIHAN BAYRAMOGLU[2]

Abstract. This paper presents derivations of asymptotic and exact distributions of bivariate exceedance statistics defined for some general threshold models. The asymptotic distribution of exceedance statistics appear to be marginal free, depending only on copulas of underlying joint distributions. In particular cases, the exact formulas for these new distributions can be obtained. Applications in medicine and economics are also discussed.

Keywords: order statistics, copula, bivariate distributions, exceedance statistics, random threshold.

AMS Subject Classification: 62G30, 62H05.

1. Introduction

In statistical theory and its applications an important problem is to determine the exact and asymptotic distributions of exceedance statistics, defined as the number of new observations of control samples falling into the random intervals formed by some functions of existing observations, called training samples. If $X_1, X_2, ..., X_n$ are observed values of random variable X with distribution function F, then the number of new observations $X_{n+1}, X_{n+2}, ..., X_{n+m}$ from random variable X falling into $(h_1(X_1, X_2, ..., X_n), h_2(X_1, X_2, ..., X_n))$ for some real valued functions $h_1(x_1, x_2, ..., x_n) \leq h_2(x_1, x_2, ..., x_n)$ $\forall(x_1, x_2, ..., x_n) \in \mathbb{R}^n$ are relevant to many important applications, for example in prediction of future observations and testing of hypothesis. In some cases, including $h_1(x_1, x_2, ..., x_n) = -\infty$ and $h_2(x_1, x_2, ..., x_n) = x_{(i)}$, or $h_1(x_1, x_2, ..., x_n) = x_{(i)}$ and $h_2(x_1, x_2, ..., x_n) = \infty$, where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, the probability $P\{X_{n+1}, X_{n+2}, ..., X_{n+m} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\}$ is independent from the distribution of control and training samples, which allows the identification of exact and asymptotic distributions of exceedance statistics defined for some threshold models. The distributions of these statistics are independent of distributions of initial observations, i.e. they are distribution free with respect to a class of distributions containing F. In statistical literature, there appeared numerous papers investigating structure of random intervals constructed from control samples and distributions of exceedance statistics designed for different threshold models under various conditions on initial distributions (see [2, 3, 21, 22, 25]). The exceedance statistics connected with ordered statistical data can be used in many important applications such as hypothesis testing in two sample problem, in characterizing equidistributions and in prediction of future events using previous observations. Many important applications of exceedance statistics based on records appear in hydrology (see [9-11]), in connection with investigations and

[1]Department of Basic Sciences & Humanities, Faculty of Arts & Sciences, Cyprus International University Haspolat, Nicosia, North Cyprus, Turkey

[2]Department of Mathematics, Izmir University of Economics, Izmir, Turkey

e-mail:aysegulerem@gmail.com, ismihan.bayramoglu@ieu.edu.tr

Manuscript received August 2017.

prediction of extreme events, such as flooding and extreme precipitation. There are numerous papers devoted to exceedance statistics for univariate random sequences, see [1, 3-7, 13, 19, 27]. Note that, random thresholds are also used in insurance and risk processes including the ruin thresholds, (see [24]). However, there are far fewer papers devoted to exceedances in bivariate sequences. A recent study belongs to [23], who investigated the joint distribution of sample ranks of bivariate order statistics in connection with exceedance statistics. [17] studied the finite and asymptotic distributions of excedance statistics defined for some threshold models in bivariate sequences. Considering bivariate training random sample and bivariate control sample, [17] investigated exceedance statistics, defined as the number of new observations falling into the random interval defined by previous observations. The finite and asymptotic distributions of these exceedance statistics were derived, and their applications in medicine and in ecology models related to air pollution were discussed. Recently, [18] proposed a test for investigating equality of two copula functions based on bivariate exceedance statistics.

In this paper, we consider bivariate exceedance models based on random thresholds. The exact and asymptotic distributions of considered exceedance statistics are derived, and their applications in medicine and economics are discussed.

The paper is organised as follows. In Section 2, we derive the exact and asymptotic distributions of some exceedance statistics based on random thresholds. Since the distributions are marginal free and depend only on copulas, we provide examples using different copulas including independent, Gumbel-Barnett, Ali-Mikhail-Haq and Farlie-Gumbel-Morgenstern copulas. In Section 3, we consider bivariate threshold models based on concomitants of order statistics. The exact and asymptotic distributions of exceedance statistics for considered threshold models are derived. Some examples are given, applications in medicine and economics are discussed. Lastly, Section 4 concludes the paper.

2. Exceedance statistics in the general threshold model

Let (X,Y) be an absolutely continuous bivariate random vector with survival function $\overline{F}(x,y) = \widehat{C}_1(\overline{F}_X(x),\overline{F}_Y(y))$, where $\widehat{C}_1(u,v),(u,v)\in[0,1]^2$, is a survival copula (see [26], p. 32) and $\overline{F}_X(x)$ and $\overline{F}_Y(y)$ are the marginal survival functions of X and Y, respectively. Furthermore, let $(X_1,Y_1)\,,(X_2,Y_2)\,,...,\,(X_m,Y_m)\,,...$ be a sequence of independent random vectors with survival function $\overline{G}(x,y)=\widehat{C}_2(\overline{F}_X(x),\overline{F}_Y(y))$ and the same marginal survival functions $\overline{F}_X(x)$ and $\overline{F}_Y(y)$. Here, (X,Y) and $(X_1,Y_1)\,,(X_2,Y_2)\,,...,\,(X_m,Y_m)\,,...$ are assumed to be independent. Let, $F_X(x)$ and $F_Y(y)$ be marginal distributions of X and Y, respectively . While $C_1(F_X(x),F_Y(y))$ is the connected copula of the random variables X and Y, $C_2(F_X(x),F_Y(y))$ is the connected copula of $(X_1,Y_1)\,,(X_2,Y_2)\,,...,\,(X_m,Y_m)\,,....$

Define the binary random variables

$$\nu_i\equiv I\left(X_i,Y_i\right)=\begin{cases}1, & X_i<X \text{ or } Y_i<Y\\ 0, & otherwise\end{cases},$$

$$\eta_i=1-\nu_i=I_M\left(X_i,Y_i\right)=\begin{cases}1, & (X_i,Y_i)\in M\\ 0, & otherwise\end{cases},$$

where $I_M\left(X_i,Y_i\right)$ is an indicator function of the random set $M\equiv M(X,Y)=[X,\infty)\times[Y,\infty)$.

Now define the random variable $S_m=\sum_{i=1}^m \nu_i$. It is clear that the exceedance statistic S_m counts the number of bivariate observations (X_i,Y_i), $i=1,2,...$ falling into the random set M^c. Let us note that the random variables ν_i, $i=1,2,...$ are dependent. In Theorem 2.1, we obtain the finite distribution of S_m.

Theorem 2.1. *It is true that*

$$P\left\{S_m = k\right\} = \binom{m}{k} \int_0^1 \int_0^1 \left[1 - \widehat{C}_2(1-u, 1-v)\right]^k \left[\widehat{C}_2(1-u, 1-v)\right]^{m-k} dC_1(u,v). \tag{1}$$

and

$$P\left\{S_m = k\right\} = \binom{m}{k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[1 - \overline{G}(x,y)\right]^k \left[\overline{G}(x,y)\right]^{m-k} dF_{X,Y}\left(x,y\right).$$

Proof. First we show that

$$P\left\{S_m = k\right\} = \binom{m}{k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[1 - \overline{G}(x,y)\right]^k \left[\overline{G}(x,y)\right]^{m-k} dF_{X,Y}\left(x,y\right).$$

Indeed, denote by

$$A_{i_j} = \left\{X_{i_j} < X, Y_{i_j} < Y\right\} \cup \left\{X_{i_j} < X, Y_{i_j} > Y\right\} \cup \left\{X_{i_j} > X, Y_{i_j} < Y\right\}$$

and observe that the complement of A_{i_j} is

$$A_{i_j}^c = \left\{X_{i_j} > X, Y_{i_j} > Y\right\}.$$

For simplicity denote by

$$E_k = \left\{A_{i_1} A_{i_2} ... A_{i_k} A_{i_{k+1}}^c ... A_{i_m}^c\right\}.$$

Then by conditioning on $X = x$ and $Y = y$, we have

$$\begin{aligned} P\left\{S_m = k\right\} &= \sum_{i_1, i_2, ..., i_m} P\left\{A_{i_1} A_{i_2} ... A_{i_k} A_{i_{k+1}}^c ... A_{i_m}^c\right\} \\ &= \sum_{i_1, i_2, ..., i_m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P\left\{E_k \mid X = x, Y = y\right\} dF_{X,Y}(x,y) \\ &= \sum_{i_1, i_2, ..., i_m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P\left\{a_{i_1} a_{i_2} ... a_{i_k} a_{i_{k+1}}^c ... a_{i_m}^c\right\} dF_{X,Y}(x,y), \end{aligned} \tag{2}$$

where

$$a_{i_j} = \left\{X_{i_j} \le x, Y_{i_j} \le y\right\} \cup \left\{X_{i_j} \le x, Y_{i_j} > y\right\} \cup \left\{X_{i_j} > x, Y_{i_j} \le y\right\}.$$

$a_{i_j}^c$ is the complement of a_{i_j} and the sum $\sum_{i_1, i_2, ..., i_m}$ extends over all permutations of $i_1, i_2, ..., i_m \in \{1, 2, ..., m\}$. Since X and Y are random variables and the events A_{i_j} are dependent, by conditioning on $X = x$ and $Y = y$ we can obtain independent events a_{i_j}. It is obvious that $p\left(a_{i_1}\right) = p\left(a_{i_2}\right) = ... = p\left(a_{i_k}\right) = 1 - \overline{G}(x,y)$ and similarly $p\left(a_{i_{k+1}}^c\right) = p\left(a_{i_{k+2}}^c\right) = ... = p\left(a_{i_m}^c\right) = \overline{G}(x,y)$. By using the independency of events a_{i_j}, the probability in (2) is

$$\begin{aligned} &P\left\{a_{i_1} a_{i_2} ... a_{i_k} a_{i_{k+1}}^c ... a_{i_m}^c\right\} \\ &= P(a_{i_1}) P(a_{i_2}) \cdots P(a_{i_k}) P(a_{i_{k+1}}^c) ... P(a_{i_m}^c). \end{aligned}$$

Therefore,

$$P\left\{S_m = k\right\} = \binom{m}{k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[1 - \overline{G}(x,y)\right]^k \left[\overline{G}(x,y)\right]^{m-k} dF_{X,Y}\left(x,y\right).$$

Using the probability integral transformation $F_X(x) = u$ and $F_Y(y) = v$, we obtain (1). $\square$

Corollary 2.1. *It is clear that*

$$E\left(S_m\right) = m - mE\left(\overline{G}(X,Y)\right),$$
$$Var(S_m) = m\left[E(\overline{G}(X,Y)) - E(\overline{G}(X,Y)^2)\right] + m^2 Var\left(\overline{G}\left(X,Y\right)\right).$$

Proof. In fact, let $h\left(x,y\right)$ be any integrable function, then

$$E\left(h\left(X,Y\right)\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h\left(x,y\right) dF\left(x,y\right).$$

Therefore, for

$$h\left(x,y\right) = \sum_{k=0}^{m} k\binom{m}{k}\left[1-\overline{G}(x,y)\right]^k\left[\overline{G}(x,y)\right]^{m-k}$$

and

$$E\left(S_m\right) = \sum_{k=0}^{m} k\binom{m}{k}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[1-\overline{G}(x,y)\right]^k\left[\overline{G}(x,y)\right]^{m-k} dF\left(x,y\right)$$

one can write

$$\begin{aligned}
E\left(S_m\right) &= E\left(\sum_{k=0}^{m} k\binom{m}{k}\left[1-\overline{G}(X,Y)\right]^k\left[\overline{G}(X,Y)\right]^{m-k}\right)\\
&= E\left(\sum_{k=0}^{m} k\frac{m!}{k!\left(m-k\right)!}\left[1-\overline{G}(X,Y)\right]^k\left[\overline{G}(X,Y)\right]^{m-k}\right)\\
&= E\left(m\left(1-\overline{G}(X,Y)\right)\sum_{i=0}^{m-1}\binom{m-1}{i}\left[1-\overline{G}(X,Y)\right]^i\left[\overline{G}(X,Y)\right]^{m-1-i}\right)\\
&= E\left(m\left(1-\overline{G}(X,Y)\right)\right) = m - mE\left(\overline{G}(X,Y)\right). \qquad (3)
\end{aligned}$$

Similarly, the second moment of S_m is

$$\begin{aligned}
E\left(S_m^2\right) &= \sum_{k=0}^{m} k^2\binom{m}{k}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[1-\overline{G}(x,y)\right]^k\left[\overline{G}(x,y)\right]^{m-k} dF\left(x,y\right)\\
&= E\left(\sum_{k=0}^{m} k^2\binom{m}{k}\left[1-\overline{G}(X,Y)\right]^k\left[\overline{G}(X,Y)\right]^{m-k}\right)\\
&= m - mE(\overline{G}(X,Y)) - mE\left(\left(1-\overline{G}(X,Y)\right)^2\right) + m^2E\left(\left(1-\overline{G}(X,Y)\right)^2\right). \qquad (4)
\end{aligned}$$

Then

$$\begin{aligned}
Var\left(S_m\right) &= E\left(S_m^2\right) - \left(E\left(S_m\right)\right)^2\\
&= m\left[E(\overline{G}(X,Y)) - E(\overline{G}(X,Y)^2)\right] + m^2Var\left(\overline{G}\left(X,Y\right)\right). \qquad (5)
\end{aligned}$$

□

Example 2.1. Consider the trivial case where the random variables X and Y are independent and so are X_i and Y_i. Then $C_1(u,v) = C_2(u,v) = uv$.

$$P\{S_m = k\} = \binom{m}{k} \int_0^1 \int_0^1 [1-(1-u)(1-v)]^k [(1-u)(1-v)]^{m-k} \, dudv.$$

$$= \binom{m}{k} \int_0^1 \int_0^1 \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} [(1-u)(1-v)]^{m-i} \, dudv$$

$$P\{S_m = k\} = \binom{m}{k} \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \frac{1}{(m-i+1)^2}. \tag{6}$$

Below in Table 1, we provide some numerical values for the probability $P\{S_m = k\}$ calculated from (6).

Table 1. The values of $P\{S_m = k\}$ in Example 2.1.

$m\backslash k$	0	1	2	3	4	5	6	7	8	9	10
5	0.028	0.061	0.103	0.158	0.242	0.408					
10	0.009	0.017	0.028	0.038	0.05	0.068	0.085	0.11	0.14	0.180	0.275

Example 2.2. Let $\widehat{C}_2(u,v) = uv\exp(-\theta \ln u \ln v)$, $\theta \in (0,1]$ be a Gumbel-Barnett family of copulas (see [26]) and $C_1(u,v) = uv$. For this case the distribution of S_m is

$$P\{S_m = k\}$$
$$= \binom{m}{k} \int_0^1 \int_0^1 \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} [(1-u)(1-v)\exp(-\theta \ln(1-u)\ln(1-v)]^{m-i} \, dudv.$$

Some numerical values of $P\{S_m = k\}$ for different values of θ are provided below in Table 2.

Table 2. The values of $P\{S_m = k\}$ in Example 2.2.

	$m\backslash k$	0	1	2	3	4	5	6	7	8	9	10
$\theta = 0.3$	5	0.0267	0.0577	0.0955	0.1458	0.2256	0.4487					
$\theta = 0.5$	5	0.026	0.056	0.091	0.139	0.217	0.471					
$\theta = 1$	5	0.025	0.052	0.083	0.13	0.20	0.51					
$\theta = 1$	10	0.01	0.02	0.024	0.033	0.042	0.054	0.067	0.08	0.11	0.16	0.40

Example 2.3. Let $C_1(u,v) = uv$ and $\widehat{C}_2(u,v) = \frac{uv}{1-\theta(1-u)(1-v)}$, $\theta \in [-1,1)$ be the Ali-Mikhail-Haq family of copulas, (see [8]). Then

$$P\{S_m = k\} = \binom{m}{k} \int_0^1 \int_0^1 \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \left(\frac{(1-u)(1-v)}{1-\theta uv} \right)^{m-i} dudv.$$

Some numerical values of $P\{S_m = k\}$ for $\theta = -1$ are given below in Table 3:

Table 3. The values of $P\{S_m = k\}$ in Example 2.3.

$m\backslash k$	0	1	2	3	4	5	6	7	8	9	10
5	0.03	0.06	0.09	0.14	0.23	0.45					
10	0.01	0.016	0.025	0.03	0.046	0.059	0.076	0.098	0.13	0.19	0.32

2.1. **Asymptotic distribution of exceedance statistic based on order statistics.** Now consider

$$\nu_i^{(m)} = \begin{cases} 1, & X_{i:m} < X \text{ or } Y_{i:m} < Y \\ 0, & otherwise \end{cases},$$

$$\eta_i^{(m)} = 1 - \nu_i^{(m)} = I_M\left(X_{i:m}, Y_{i:m}\right) = \begin{cases} 1, & \left(X_{i:m}, Y_{i:m}\right) \in M \\ 0, & otherwise \end{cases},$$

where $X_{i:m}$ and $Y_{i:m}$ are the ith order statistics of $X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_m$, respectively. Furthermore $I_M\left(X_{i:m}, Y_{i:m}\right)$ is an indicator function of the random set $M \equiv M(X,Y) = [X, \infty) \times [Y, \infty)$. Define the random variable $S_m' = \sum\limits_{i=1}^{m} \nu_i^{(m)}$. In Theorem 2.2, the asymptotic distribution of $\frac{S_m'}{m}$ is derived.

Theorem 2.2. *The asymptotic distribution of* $\frac{S_m'}{m}$ *is*

$$\lim_{m\to\infty} \sup_{-\infty<x<\infty} \left| P\left\{\frac{S_m'}{m} \le x\right\} - P\{U + V - C_2(U,V) \le x\} \right| = 0. \tag{7}$$

or

$$\lim_{m\to\infty} \sup_{-\infty<x<\infty} \left| P\left\{\frac{S_m'}{m} \le x\right\} - P\left\{F_X(X) + F_Y(Y) - G(X,Y) \le x\right\} \right| = 0.$$

where (U,V) *is a bivariate random vector having Uniform (0,1) marginals and connecting copula* $C_2(u,v)$.

Note. It is clear that (7) can also be written as

$$\lim_{m\to\infty} \sup_{-\infty<x<\infty} \left| P\left\{\frac{S_m'}{m} \le x\right\} - P\{1 - \widehat{C}_2(1-U, 1-V) \le x\} \right| = 0.$$

Proof. Let

$$G_m^*\left(u,v\right) = \frac{1}{m} \sum_{i=1}^{m} I_{(-\infty,u]\times(-\infty,v]}\left(X_i, Y_i\right)$$

be a bivariate empirical distribution of the sample $(X_1, Y_1), ..., (X_m, Y_m)$. It can be easily shown for any integrable function $g(x,y)$ that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g\left(u,v\right) dG_m^*\left(u,v\right) = \sum_{i=1}^{m} g\left(X_{i:m}, Y_{i:m}\right).$$

Consider

$$
\begin{aligned}
P\left\{\frac{S_m^{'}}{m} \leq x\right\} &= P\left\{\frac{1}{m}\sum_{i=1}^{m}\nu_i^{(m)} \leq x\right\} \\
&= P\left\{\frac{1}{m}\sum_{i=1}^{m}(1-\eta_i^{(m)}) \leq x\right\} && (8)\\
&= P\left\{1-\frac{1}{m}\sum_{i=1}^{m}\eta_i^{(m)} \leq x\right\} && (9)\\
&= 1-P\left\{\frac{1}{m}\sum_{i=1}^{m}\eta_i^{(m)} \leq 1-x\right\} && (10)\\
&= 1-P\left\{\frac{1}{m}\sum_{i=1}^{m}I_{[X,\infty)\times[Y,\infty]}\left(X_{i:m},Y_{i:m}\right) \leq 1-x\right\} \\
&= 1-P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{X,Y}\left(u,v\right)dG_m^{*}\left(u,v\right) \leq 1-x\right\}, && (11)
\end{aligned}
$$

where

$$g_{X,Y}(u,v) = I_{[X,\infty]\times[Y,\infty)}\left(u,v\right).$$

Conditioning on $X = x$ and $Y = y$ and using independence of random vectors (X,Y) and (X_i,Y_i), $i = 1,2,...,m$, we obtain

$$
\begin{aligned}
&P\left\{\frac{S_m^{'}}{m} \leq x\right\} \\
&= 1-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{X,Y}\left(u,v\right)dG_m^{*}\left(u,v\right) \leq 1-x \mid X=x, Y=y\right\} dF(x,y) \\
&= 1-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{x,y}\left(u,v\right)dG_m^{*}\left(u,v\right) \leq 1-x\right\} dF(x,y),
\end{aligned}
$$

and

$$G_m^{*}\left(u,v\right) = \frac{1}{m}\sum_{i=1}^{m}I_{(-\infty,u]\times(-\infty,v]}(X_i,Y_i)$$

is the empirical distribution function constructed by the first m observations (X_i,Y_i), $i = 1,2,...,m$. From the Glivenko Cantelli Theorem (see [12], page 5)

$$\sup_{(u,v)\in\mathbb{R}^2} \left|G_m^{*}\left(u,v\right) - G(u,v)\right| \overset{a.s.}{\to} 0, \text{ as } m \to \infty.$$

using the continuity property of the integral functional

$$\Psi(G) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{x,y}\left(u,v\right)dG\left(u,v\right)$$

we have

$$\Psi(G_m^{*}) \overset{a.s.}{\to} \Psi(G).$$

Then we can write

$$
\begin{aligned}
&\lim_{m\to\infty} P\left\{\frac{S'_m}{m}\le x\right\}\\
&= 1-\lim_{m\to\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{x,y}(u,v)\,dG_m^*(u,v)\le 1-x\right\}dF(x,y)\\
&= 1-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{x,y}(u,v)\,dG(u,v)\le 1-x\right\}dF(x,y)\\
&= 1-P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{X,Y}(u,v)\,dG(u,v)\le 1-x\right\}\\
&= 1-P\left\{\int_{X}^{\infty}\int_{Y}^{\infty} dG(u,v)\le 1-x\right\}\\
&= 1-P\{\overline{G}(X,Y)\le 1-x\}. \qquad (12)\\
&= P\left\{F_X(X)+F_Y(Y)-G(X,Y)\le x\right\}. \qquad (13)
\end{aligned}
$$

Finally, using probability integral transformation $U=F_X(X)$ and $V=F_Y(Y)$ in (13) we obtain (7). Thus, the theorem is proved. □

Corollary 2.2. *If $C_1(u,v)=C_2(u,v)=C(u,v)$, then*

$$\lim_{m\to\infty}\sup_{-\infty<x<\infty}\left|P\left\{\frac{S'_m}{m}\le x\right\}-P\{U+V-C(U,V)\le x\}\right|=0 \qquad (14)$$

and

$$\lim_{m\to\infty}\sup_{-\infty<x<\infty}\left|P\left\{\frac{S'_m}{m}\le x\right\}-P\{1-\widehat{C}(1-U,1-V)\le x\}\right|=0,$$

where (U,V) has the copula $C(u,v)$.

Example 2.4. Let $C_1(u,v)=C_2(u,v)=uv$. Then the asymptotic distribution of $T=\frac{S'_m}{m}$ is

$$
\begin{aligned}
F_T(x) &\equiv \lim_{m\to\infty} P\left\{\frac{S'_m}{m}\le x\right\}\\
&= P\{U+V-C(U,V)\le x\} = \iint_{\{(u,v):\ u+v-uv\le x\}} du dv.
\end{aligned}
$$

After transformations $t=1-(1-u)(1-v)$ and $s=1-u$ with Jacobian $J=\left|\frac{1}{s}\right|$, we have

$$F_T(x)=\iint_{\left\{\begin{array}{c}(t,s):\ 0\le t\le x\\ 1-t\le s\le 1\end{array}\right\}} \frac{1}{s}dtds,$$

and

$$
\begin{aligned}
&F_T(x)=(1-x)\ln(1-x)+x,\ 0\le x\le 1, \qquad (15)\\
&f_T(x)=-\ln(1-x),\ 0\le x\le 1,
\end{aligned}
$$

where $f_T(x) = \frac{dF_T(x)}{dx}$ is a pdf of the limiting distribution. Below in Figure 1, the graphs of the functions $F_T(x)$ and $f_T(x)$ are illustrated.

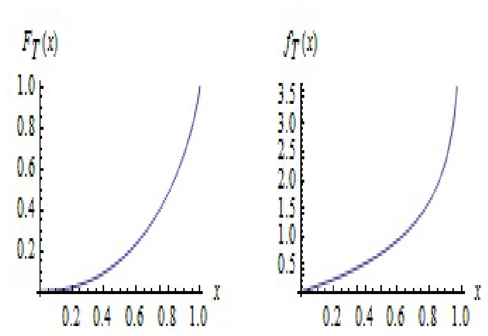


Figure 1. The graphs of $F_T(x)$ and $f_T(x)$.

Example 2.5. Let $C_1(u,v) = uv[1+\alpha(1-u)(1-v)]$, $-1 \le \alpha \le 1$, be the Farlie-Gumbel-Morgenstern distribution and $C_2(u,v) = uv$. Then

$$\lim_{m\to\infty} P\left\{\frac{S'_m}{m} \le x\right\} = P\{U+V-C_2(U,V) \le x\} = \iint_{\{(u,v):\ u+v-uv\le x\}} (1+\alpha(1-2u)(1-2v))dudv$$

We denote by $W_\alpha = \frac{S'_m}{m}$ in Example 5. After making similar transformations as in example 4, we have

$$\begin{aligned} F_{W_\alpha}(x) &\equiv \lim_{m\to\infty} P\left\{\frac{S'_m}{m} \le x\right\} \\ &= (1-3\alpha(x-1))\,x + (x-1)\,(\alpha\,(2x-3)-1)\ln(1-x)\,, \\ &-1 \le \alpha \le 1 \text{ and } 0 \le x \le 1. \end{aligned}$$

$$\begin{aligned} f_{W_\alpha}(x) &= \frac{F_{W_\alpha}(x)}{dx} \\ &= (\alpha(4x-5)-1)\ln(1-x) - 4\alpha x,\ 0 \le x \le 1. \end{aligned}$$

Below in Table 4, we provide some numerical values for the moments and skewness of W_α. Then, it is observed that, if α increases from -1 to 1, skewness also increases.

Table 4. Moments and variance of W_α for selected values of α.

α	$E(W_\alpha)$	$E\left(W_\alpha^2\right)$	$E\left(W_\alpha^3\right)$	$Var\,(W_\alpha)$	$Skewness$
-0.9	0.775	0.6361	0.5411	0.0355	-1.0331
-0.5	0.7639	0.625	0.5321	0.0415	-1.0312
-0.1	0.7528	0.6139	0.5231	0.0472	-0.9867
0.1	0.7472	0.6083	0.5186	0.04999	-0.9561
0.5	0.7361	0.5972	0.5096	0.0554	-0.8863
0.9	0.725	0.5861	0.5006	0.0605	-0.8102

Below in Figure 2, the graphs of $F_{W_\alpha}(x)$ and $f_{W_\alpha}(x)$ with respect to α for fixed values of x are given. The graphs show that larger values of α corresponds to larger values of $F_{W_\alpha}(x)$ for fixed x. The pdf $f_{W_\alpha}(x)$ varies depending on values of both α and x.

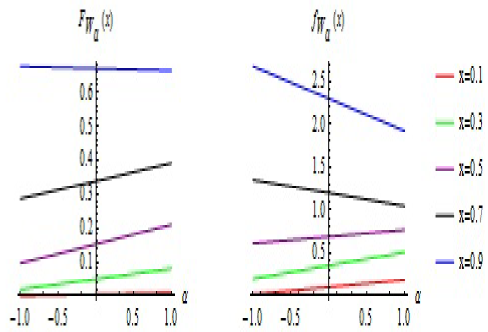


Figure 2. The graphs of $F_{W\alpha(x)}$ and $f_{W\alpha}(x)$.

3. Random thresholds based on order statistics and concomitants

Let $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ be a sequence of independent random vectors with joint distribution function $F(x, y) = C_1(F_X(x), F_Y(y))$ and let $(X_{n+1}, Y_{n+1}), (X_{n+2}, Y_{n+2}), ...,$ $(X_{n+m}, Y_{n+m}), ...$ be another sequence of independent random vectors with joint survival function $\bar{G}(x, y) = \widehat{C}_2(F_X(x), F_Y(y))$. We assume that the random vectors $(X_1, Y_1), (X_2, Y_2), ...,$ (X_n, Y_n) and the sequence of random vectors $(X_{n+1}, Y_{n+1}), ..., (X_{n+m}, Y_{n+m}), ...$ are independent.

Let $(X_{r:n}, Y_{[r:n]})$ be the vector of rth order statistic and its concomitant constructed from the sample $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$. The theory of concomitants is well documented in [14-16, 20]. The joint pdf of the rth order statistic $X_{r:n}$ and its concomitant $Y_{[r:n]}$ is (see [16])

$$f_{X_{r:n}, Y_{[r:n]}}(x, y) = f(y \mid x) f_{r:n}(x). \tag{16}$$

Denote $X_{n+i} = X_i^{'}$ and $Y_{n+i} = Y_i^{'}$. Then $X_{i:m}^{'}$ and $Y_{i:m}^{'}$ are the ith order statistics of $X_1^{'}, X_2^{'}, ..., X_m^{'}$ and $Y_1^{'}, Y_2^{'}, ..., Y_m^{'}$, respectively.

Define the binary random variables

$$\nu_i(r) \equiv I(X_i, Y_i) = \begin{cases} 1, & X_i^{'} < X_{r:n} \text{ or } Y_i^{'} < Y_{[r:n]} \\ 0, & otherwise \end{cases},$$

$$\eta_i(r) = 1 - \nu_i(r) = I_N\left(X_i^{'}, Y_i^{'}\right) = \begin{cases} 1, & (X_i^{'}, Y_i^{'}) \in N \\ 0, & otherwise \end{cases},$$

where $N \equiv N(X_{r:n}, Y_{[r:n]}) \equiv [X_{r:n}, \infty) \times [Y_{[r:n]}, \infty)$. Now define the random variable $S_m(r) = \sum_{i=1}^{m} \nu_i(r)$.

Theorem 3.1. *It is true that*

$$P\{S_m(r) = k\} = \binom{m}{k} \frac{1}{Beta(r, n-r+1)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x)]^{r-1} \times$$
$$[1 - F(x)]^{n-r} [\overline{G}(x, y)]^{m-k} [1 - \overline{G}(x, y)]^k dF(x, y) \tag{17}$$

and

$$P\{S_m(r) = k\} = \binom{m}{k} \frac{1}{Beta(r, n-r+1)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u^{r-1} (1-u)^{n-r} [\widehat{C}_2(1-u, 1-v)]^{m-k} \times$$

$$[1-\widehat{C}_2(1-u,1-v)]^k dC_1(u,v). \tag{18}$$

Proof. Proof is similar to proof of Theorem 2.1. □

Corollary 3.1. *It is clear that*

$$\begin{aligned} E(S_m(r)) &= m - mE(\overline{G}(X_{r:n},Y_{[r:n]})), \\ Var(S_m(r)) &= mE(\overline{G}\left(X_{r:n},Y_{[r:n]}\right) \\ &\quad - E(\overline{G}\left(X_{r:n},Y_{[r:n]}\right)^2)) + m^2 Var\left(\overline{G}\left(X_{r:n},Y_{[r:n]}\right)\right). \end{aligned}$$

Proof. Proof is similar to proof of Corollary 2.1. □

Example 3.1. Consider $C_1(u,v)=C_2(u,v)=uv$. Then,

$$P\{S_m(r)=k\} = \binom{m}{k}\frac{1}{Beta(r,n-r+1)}\sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}\frac{Beta(r,n+m-r-i+1)}{m-i+1}.$$

Below in Table 5, we provide some numerical values of $P\{S_m(r)=k\}$ for different values of m,n, and r.

Table 5. Some numerical values of $P\{S_m(r)=k$ for different values of m,n and r.

k	0	1	2	3	4	5	6	7	8	9	10
$m=5/n=4/r=3$	0.008	0.032	0.08	0.16	0.28	0.44					
$m=10/n=4/r=4$	0.00009	0.0004	0.001	0.004	0.0091	0.018	0.03241	0.065	0.12	0.23	0.52

3.1. **Asymptotic distributions of exceedance statistics based on order statistics.** Let $X'_{i:m}$ and $Y'_{i:m}$ are the ith order statistics of $X'_1,X'_2,...,X'_m$ and $Y'_1,Y'_2,...,Y'_m$, respectively.

Now consider

$$\nu_i^{(m)}(r) = \begin{cases} 1, & X'_{i:m} < X_{r:n} \text{ or } Y'_{i:m} < Y_{[r:n]} \\ 0, & otherwise \end{cases},$$

$$\eta_i^{(m)}(r) = 1-\nu_i^{(m)}(r) = I_N\left(X'_{i:m},Y'_{i:m}\right) = \begin{cases} 1, & (X'_{i:m},Y'_{i:m}) \in N \\ 0, & otherwise \end{cases},$$

where $N \equiv N(X_{r:n},Y_{[r:n]}) \equiv [X_{r:n},\infty)\times[Y_{[r:n]},\infty)$. Now define the random variable $S'_m(r) = \sum_{i=1}^m \nu_i^{(m)}(r)$. In Theorem 3.2 the asymptotic distribution of $\frac{S'_m}{m}$ is given.

Theorem 3.2. *The asymptotic distribution of* $\frac{S'_m(r)}{m}$ *is*

$$\lim_{m\to\infty}\sup_{-\infty<x<\infty}\left|P\left\{\frac{S'_m(r)}{m}\le x\right\} - P\left\{F_X(X_{r:n})+F_Y(Y_{[r:n]})-G_{X,Y}(X_{r:n},Y_{[r:n]})\le x\right\}\right| = 0. \tag{19}$$

It is clear that (19) can also be written as

$$\lim_{m\to\infty}\sup_{-\infty<x<\infty}\left|P\left\{\frac{S'_m(r)}{m}\le x\right\} - P\left\{U_{r:n}+V_{[r:n]}-C_2(U_{r:n},V_{[r:n]})\le x\right\}\right| = 0$$

and

$$\lim_{m\to\infty}\sup_{-\infty<x<\infty}\left|P\left\{\frac{S'_m(r)}{m}\le x\right\} - P\left\{1-\widehat{C}_2(1-U_{r:n},1-V_{[r:n]})\le x\right\}\right| = 0.$$

Proof. Proof is similar to proof of Theorem 2.2. □

Corollary 3.2. *Let* $C_1(u,v) = C_2(u,v) = C(u,v)$, *then*

$$\lim_{m\to\infty} \sup_{-\infty<x<\infty} \left| P\left\{ \frac{S'_m(r)}{m} \le x \right\} - \frac{1}{Beta(r,n-r+1)} \iint_{\{(u,v):u+v-C(u,v)\le x\}} u^{r-1}(1-u)^{n-r} c(u,v)dudv \right| = 0.$$

Example 3.2. Let $C_1(u,v) = C_2(u,v) = uv$, then

$$T_1(x) \equiv \lim_{m\to\infty} P\left\{ \frac{S'_m(r)}{m} \le x \right\} = \frac{n}{n-r} \int_0^x F_{X_{r:n-1}}(t)dt,$$

and

$$t_1(x) = \frac{dT_1(x)}{dx} = \frac{n}{n-r} F_{X_{r:n-1}}(x), \quad 0 < x < 1,$$

where $F_{X_{r:n-1}}(x)$ is the distribution function of the rth order statistics of $Uniform(0,1)$ distribution. The graphs of the limiting distribution $T_1(x)$ for $n = 5$, $r = 1, ..., 5$ is presented in Figure 3.

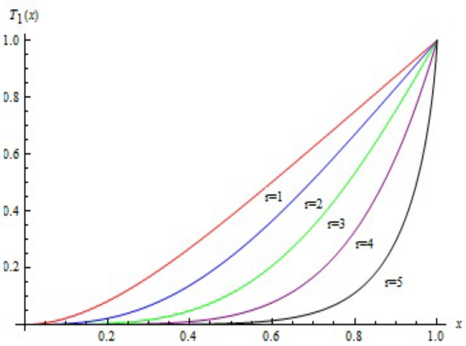


Figure 3. The graph of $T_1(x)$ for $n = 5$ and $r = 1, ..., 5$.

A note on applications

1. Application in medicine

The control of cholesterol level occupies a significant place in health policy of the developing countries since it is an important cause of heart and coroner disease, and more importantly premature death. As is well known, cholesterol is a fat that is non soluble in water, but is transported in the blood by a naturally-occuring water-soluble cholesterol called hypoproteins. The Low Density Protein (LDL) cholesterol is regarded as the most important cholesterol type for detecting and treating high cholesterol. However, high values of LDL can cause series heart problems. In particular, high LDL cholesterol levels, combined with high triglycerides levels raise the risk of heart problems and stroke. Therefore, high cholesterol levels known as the "silent killer". HDL is known to be positive (good) when high. A variety of factors are related with high cholesterol, such as nutritional habits, weight, physical activity, age, and heredity. Three cholesterol levels which are defined in [28, 29] are given in Table 6.

Table 6. Cholesterol levels.

Category	Desirable	Borderline	Risky
Total Cholesterol	<200 mg/dL	200-239 mg/dL	$\geq$ 240 mg/dL
LDL	<130 mg/dL	130-159 mg/dL	$\geq$ 160 mg/dL
Triglycerides	<150 mg/dL	150-199 mg/dL	$\geq$ 200 mg/dL
HDL	>60 mg/dL	40-60 mg/dL	$\leq$ 40 mg/dL

Assume that, given fixed good HDL level (say HDL$>$ 60 mg/dL) and high total cholesterol (say Total Cholesterol Level $\geq$ 240 mg/dL) m patients are tested for risk of cardiovascular disease. One of the contemporary approaches is that if HDL is high, then high levels of LDL in the presence of low levels of triglycerides or high levels of triglyceride in the presence of low levels of LDL may be based on genetic factors, patients are therefore less likely to risk of cardiovascular disease. Let interpret the random vectors (X_i, Y_i), $i = 1, 2, ..., n$ as a training sample of the patients with high LDL cholesterol or high triglycerides levels, where X_i denotes the LDL cholesterol level and Y_i denotes the triglycerides level of the ith patient. Then (X_{n+j}, Y_{n+j}), $j = 1, 2, ...$ be the LDL cholesterol and triglycerides levels of patients in control sequence. Then it is clear that the statistics $m - S_m(r)$ denotes the number of patients with high risk in cardiovascular disease.

2. Application in exchange rate

In financial mathematics because the prediction and modelling of the exchange rates is so challenging, interfaces about exchange rates are valuable. Let $(X_1, Y_1), (X_2, Y_2), ..., (X_{m,}Y_m)$ be the exchange rate at the end of the each day, where (X_i, Y_i) denotes the exchange rate of Dollar and Euro at the end of ith day, respectively. Consider the bivariate random variables (X, Y) which denotes the critical random thresholds for USD and EUR, respectively. The random vector (X, Y) and the sequence (X_i, Y_i) are assumed to be independent. It is clear that $E(S_m)$ is the expected number of the days on which the USD or EUR will exceed the critical random threshold (X, Y). The asymptotic distribution of S'_m/m can be used in the construction of prediction intervals for the financial variables that determine long term economic stability.

4. Conclusion

In this paper, we consider some general bivariate random exceedance models based on random thresholds. The random thresholds include random vectors, bivariate order statistics and concomitants. For different models the exact and asymptotic distributions of exceedance statistics are obtained. Examples with different copulas such as independent, Gumbel Barnett, Ali-Mikhail-Haq and Farlie-Gumbel-Morgenstern copulas are provided. The finite and asymptotic distributions obtained in the paper present theoretical and practical interest. Some possible applications in medicine and economics are discussed.

5. Acknowledgement

The authors thank the anonymous reviewer and editor for their valuable comments and suggestions that led to improvements in the paper.

References

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

[2] Bairamov, I.G., Petunin, Yu.I., (1991), The structure of invariant confidence intervals containing the main distributed mass, Theory Probab. Appl., 35(1), pp.15–26.

[3] Bairamov, I.G., Khan, M.K., (2007), On exceedances of record and order statistics, Proc. Amer. Math. Soc., 135(6), pp.1935-1945.

[4] Bairamov, I., Tanil, H., (2007), On the distributions of exceedance based on generalized order statistics, Comm. Statist. Theory Methods, 36(5-8), pp.1479-1491.

[5] Bairamov, I.G., Eryilmaz, S., (2008), Joint behavior of precedences and exceedances in random threshold models, Aust. N. Z. J. Stat., 50(3), pp.209-216.

[6] Bairamov, I.G., Kotz, S., (2001), On distributions of exceedances associated with order statistics and record values for arbitrary distributions, Statist. Papers, 42(2), pp.171–185.

[7] Bayramoglu, I., Giner, G., (2014), Order statistics and exceedances for some models of INID random variables, Braz. J. Probab. Stat., 28(4), pp.492-514.

[8] Balakrishnan, N., Lai, C.D., (2009), Continuous Bivariate Distributions, Springer, NY.

[9] Benestad, R.E., (2003), How often can we expect a record event?, Clim. Res., 25, pp.3-13.

[10] Benestad, R.E., (2004), Record-values, nonstationarity tests and extreme value distributions, Glob. Planet. Chang., 44, pp.11-26.

[11] Benestad, R.E., (2006), Can we expect more extreme precipitation on the monthly time scale?, J. Clim., 19(4), pp.630-637.

[12] Borovkov, A.A., (1998), Mathematical Statistics. Gordon and Breach Science Publishers, Amsterdam.

[13] Boutsikas, M.V., Koutras M.V., (2002), Modelling claim exceedances over thresholds, Insur. Math. Econ., 30, pp.67–83.

[14] David, H.A., Nagaraja, H.N., (2003), Order Statistics, 3rd edition, John Wiley & Sons, New Jersey.

[15] David, H.A., Nagaraja, H.N., (1998), Concomitants of order statistics, In Handbook of Statistics 16: Order Statistics: Theory & Methods (Eds., N. Balakrishnan and C.R. Rao), pp.487-513, North Holland Academic Publishers, Amsterdam.

[16] David, H.A., O' Connecll, M.J.,Yang, S.S., (1977), Distribution and expected value of the rank of a concomitant of an order statistic, Ann Statist., 5, pp.216–223.

[17] Erem, A., Bayramoglu I., (2017), Exact and asymptotic disributions of exceedance statistics for bivariate random sequences, Statistics & Probability Letters, 125, pp.181–188.

[18] Erem, A., (2020), Bivariate two sample test based on exceedance statistics, Comm. Statist. Simulation Comput., 49(9), pp.2389-2401.

[19] Eryilmaz, S., Gebizlioğlu O. L., Tank, F., (2001), Modelling of claims exceedances over random thresholds for related insurance, Insur. Math. Econ., 49, pp.496-500.

[20] He, Q., Nagaraja, H.N., (2009), Distribution of concomitants of order statistics and their order statistics, J. Statist. Plann. Inference, 139(8), pp.2643-2655.

[21] Katzenbeisser, W, (1985), The distribution of two sample location exceedance test statistics under Lehmann alternatives, Statist. Hefte,, 26(2), pp.131-138.

[22] Katzenbeisser, W., (1986), The exact power of two-sample location test based on exceedance against shift alternatives, Statistics, 20(1), 20, pp.47-54.

[23] Kemalbay, G., Bayramoğlu, I., (2015), Joint distribution of new sample rank of bivariate order statistics, J. Appl. Stat., 42(10), pp.2280-2289.

[24] Kemaloglu, S.A., Gebizlioglu, O.L., (2009), Risk analysis under Progressive Type-II censoring with binomial claim numbers, J. Comput. Appl. Math., 233(1), pp.61-72.

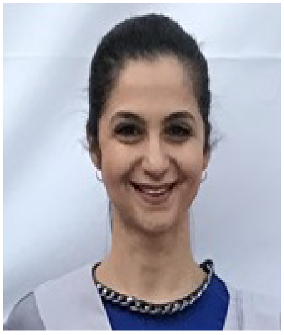
[25] Matveycuk, S.A., Petunin, Yu.I., (1990/1991), A generalization of the Bernoulli model arising in order statistics, I/II Ukr. Math. J., 42, 518-528 / 43, pp.779-785.

[26] Nelsen, R.B., (2006), An Introduction to Copulas, 2nd edition, Springer, USA.

[27] Turhan, N.O., (2012), On asymptotic distribution of exceedance statistics, J. Comput. Appl. Math., 239(9) , pp.2259-2267.

[28] Wilson, P.W.F., D'Agostino, R.B., Levy, D., Belanger, A.M, Silbershatz, H., Kannel, W.B., (1998), Prediction of coronary heart disease using risk factor categories, Circulation, 97(18), pp.1837-847.

Ismihan Bayramoglu (Ismihan Bairamov) is presently a Professor of Mathematics and Statistics and Dean of Faculty of Arts and Sciences, Izmir University of Economics, Turkey. He graduated from Faculty of Applied Mathematics of Azerbaijan State University and received his Ph.D. degree in probability and statistics from Kiev University, Ukraine in 1988. His research interests focus mainly in theory of statistics, probability, theory order statistics, reliability analysis and engineering, nonparametric statistics and multivariate distributions and copulas.



Aysegul Erem graduated from Department of Statistics of Ege University in 2012 and received her Ph.D. degree in Applied Mathematics and Statistics from Izmir University of Economics in 2017. She has been teaching in department of Basic Sciences Humanities of Cyprus International University since 2017. Her research interests focus mainly in theory of statistics, probability, theory order statistics.