https://communications.science.ankara.edu.tr

Commun.Fac.Sci.Univ.Ank.Ser. A1 Math. Stat.
Volume 70, Number 2, Pages 785–795 (2021)
DOI: 10.31801/cfsuasmas.816462
ISSN 1303–5991 E-ISSN 2618–6470

Received by the editors: October 26, 2020; Accepted: June 21, 2021

AN EXCEEDANCE MODEL BASED ON BIVARIATE ORDER STATISTICS

Aysegul EREM

Department of Basic Sciences and Humanities, Faculty of Arts and Sciences, Cyprus International University, Lefkosa, Mersin 10, TURKEY

Abstract. In hydrologic risk analysis, the use of exceedance statistics are very important. In this sense, we construct a random threshold model based on bivariate order statistics. The exact distribution of exceedance statistics is calculated under some well-known copulas such as independent and Farlie-Gumbel-Morgenstern (FGM) copulas. Furthermore, numerical results are provided for expected value and variance of exceedance statistics under independent and Farlie-Gumbel-Morgenstern copulas. The application of the model in hydrology is also discussed.

1. Introduction

Exceedance statistics and random threshold models are very useful tools in real life applications. There have been many studies about the applications of exceedances in different areas such as hydrology, actuarial sciences and medicine, see [13, 15, 12] and [19], respectively.

Eryilmaz [11], construct a random threshold model by using univariate order statistics. The distribution of the longest run statistics are derived. Then the use of the model in hydrology is discussed. For univariate random threshold models we refer to [6, 2, 21, 17, 18] and [4].

Theoretical properties and application areas of bivariate random threshold models have been discussed in many publications. In [10], marginal distribution and joint distribution of the new sample rank of rth order statistics and its concomitant are obtained. The application of the model in hydorology is discussed based on numerical results. Bayramoglu and Giner [5], construct a random threshold

2020 *Mathematics Subject Classification.* 62H05, 62E15, 62H10.

Keywords and phrases. Exceedance probability, bivariate order statistics, bivariate distributions, random threshold model, hydrologic risk.

aysegulerem@gmail.com

0000-0002-7713-5005.

model based on order statistics from independent but not necessarily identically distributed (INID) random variables. Asymptotic distributions of exceedance statistic is derived based on hypergeometric function and incomplete beta functions. Bayramoglu and Eryilmaz [4], compose a random threshold model based on two sets of exchangeable random vectors. The reliability applications of the model are discussed under the FGM distribution. In [7] and [9] bivariate random threshold models are composed based on concomitants of order statistics. Then the exact and asymptotic distribution of exceedance statistics are obtained. Applications in medicine, economics and air pollution are discussed. In [8], a statistical test is introduced for checking the equality of two copulas based on a bivariate random threshold model.

In hydrological analysis, if the flood peak and flood volume exceed critical values within a certain period, they create a risky situation. Therefore, the use of exceedance statistics and random thresholds in calculating these risk probabilities is quite important. In this study, a bivariate random threshold model based on bivariate order statistics is considered. Here we have a training sample which consists of bivariate random variables that represent flood peak and flood volume of n hydrological stations in a certain location, in the past year. We also have a bivariate control sample which consists of bivariate random variables that represent flood peak and flood volume of m hydrological stations in the same location, in the coming year. Then by using the minimum flood peak and minimum flood volume in training sample, the random threshold model is constructed. The use of the model in hydrological risk analysis is also discussed.

This paper is organized as follows: In section 2, the problem statement is provided. Then the exact distribution of exceedance statistics are obtained in terms of copula functions. Expected value and variance of exceedance statistics are provided as numerically for independent and Farlie-Gumbel -Morgenstern copulas. Lastly, Section 3 concludes the paper.

2. Model Description

Let $T_1 = \{(X_k, Y_k), k = 1, 2, ..., n\}$ be a sequence of independent random variables with joint cumulative distribution function (CDF) $F(x, y) = C_1(F_X(x), F_Y(y))$, where $C_1(u, v)$, $(u, v) \in [0, 1]^2$ is a connecting copula and $F_X(x), F_Y(y)$ are the marginal CDF' s of X and Y, respectively. Furthermore, let $T_2 = \{(X_k^{'}, Y_k^{'}), k = 1, 2, ..., m\}$ be another sequence of independent random variables with joint CDF $G(x, y) = C_2(F_X(x), F_Y(y))$, where $C_2(u, v)$, $(u, v) \in [0, 1]^2$ is a connecting copula and $F_X(x), F_Y(y)$ are the marginal CDF' s of X and Y, respectively. Let $f(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$, $g(x, y) = \frac{\partial^2 G(x,y)}{\partial x \partial y}$, $f_X(x) = \frac{dF_X(x)}{dx}$, $f_Y(y) = \frac{dF_Y(y)}{dy}$, $\overline{F}_X(x) = 1 - F_X(x)$, and $\overline{F}_Y(y) = 1 - F_Y(y)$. Here X_k and Y_k denote flood peak and flood volume of n stations in past for a certain location, respectively. Furthermore, $X_k^{'}$

and $Y_k^{'}$ denote flood peak and flood volume of the future m stations in the same location, respectively. Here we call T_1 as training sample and T_2 as control sample.

We define the rth bivariate order statistics of T_1 as $(X_{r:n}, Y_{r:n})$, where $1 \leq r \leq n$, $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ and $Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}$ are the order statistics of $\{X_k, k = 1, 2, ..., n\}$ and $\{Y_k, k = 1, 2, ..., n\}$, respectively. For $r = 1$, $(X_{1:n}, Y_{1:n})$ denotes the smallest flood peak and flood volume in the past, respectively. Then the exceedance statistic $M_m\,(1)$ is defined as follows

$$M_m\,(1) = \sum_{k=1}^{m} \delta_k,$$

where

$$\delta_k = \begin{cases} 1, & \text{if } \left(X_k^{'}, Y_k^{'}\right) \in A_1 \\ 0, & \text{otherwise} \end{cases}$$

and $A_1 = (-\infty, X_{1:n}] \times (-\infty, Y_{1:n}]$. The set A_1 is constructed from training sample T_1.

Here $M_m\,(1)$ denotes the number of nonhazardous stations in the future observations. For example, if $M_m\,(1) = 4$ it means that there can be 4 nonhazardous stations in the future observations.

In Corollary 1, the probability mass function (PMF) of $M_m\,(1)$ is given by using the distribution of bivariate order statistics, see [3] and [14]. For $1 \leq r, s \leq n$, the joint probability density function (PDF) of $X_{r:n}$ and $Y_{s:n}$ is

$$\begin{aligned} f_{X_{r:n},Y_{s:n}}(t,s) = \sum_{t_1=a_1}^{a_2} & p_1 \left[F(t,s)\right]^{t_1} \left[(F_X(t) - F(t,s))\right]^{r-1-t_1} \left[F_Y(s) - F(t,s)\right]^{s-1-t_1} \\ & \times \left[\overline{F}(t,s)\right]^{n-r-s+t_1+1} f(t,s) + \sum_{t_4=d_1}^{d_2} \sum_{t_2=c_1}^{c_2} \sum_{t_1=b_1}^{b_2} p_2 \left[F(t,s)\right]^{t_1} \\ & \times \left[(F_X(t) - F(t,s))\right]^{r-1-t_1-t_2} \left[(F_Y(s) - F(t,s))\right]^{s-1-t_1-t_4} \\ & \times \left[\overline{F}(t,s)\right]^{n-r-s+t_1+t_2+t_4} \left[F^{\cdot,1}(t,s)\right]^{t_2} \left[f_Y(s) - F^{\cdot,1}(t,s)\right]^{1-t_2} \\ & \times \left[F^{1,\cdot}(t,s)\right]^{t_4} \left[f_X(t) - F^{1,\cdot}(t,s)\right]^{1-t_4}, \end{aligned} \tag{1}$$

where $a_1 = max(0, r+s-n-1)$, $a_2 = min(r-1, s-1)$, $b_1 = max(0, r+s-n-t2-t4)$, $b_2 = min(r - t2 - 1, s - t4 - 1)$, $c_1 = max(0, r - n + 1)$, $c_2 = min(1, r - 1)$, $d_1 = max(0, s - n + 1)$, $d_2 = min(1, s - 1)$

$$\begin{aligned} \overline{F}(t,s) &= 1 - F_X(t) - F_Y(s) + F(t,s) \\ F^{1,\cdot}(t,s) &= \frac{\partial F(t,s)}{\partial t} \\ F^{\cdot,1}(t,s) &= \frac{\partial F(t,s)}{\partial s} \end{aligned}$$

and the constants p_1 and p_2 are

$$p_1 = \frac{n!}{t_1!\,(r-1-t_1)!\,(s-1-t_1)!\,(n-r-s+t_1-1)!},$$

$$p_2 = \frac{n!}{t_1!\,(r-1-t_1-t_2)!\,(s-1-t_1-t_4)!\,(n-r-s+t_1+t_2+t_4)!}.$$

Corollary 1. *The PMF of* $M_k(1)$ *is*

$$P\{M_m(1)=l\} = \binom{m}{l} G(x,y)^l (1-G(x,y))^{m-l} f_{X_{1:n},Y_{1:n}}(x,y)\,dxdy \tag{2}$$

where $f_{X_{1:n},Y_{1:n}}(x,y)$ *is the PDF of* $X_{1:n}$ *and* $Y_{1:n}$ *in training sample* T_1.

Then by using the formula of bivariate order statistics, Equation (2) can be written as follows

$$\begin{aligned} P\{M_m(1)=l\} &= \binom{m}{l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x,y)^l (1-G(x,y))^{m-l} \{n\left(\overline{F}(x,y)\right)^{n-1} \\ &\quad \times f(x,y) + n(n-1)\left(\overline{F}(x,y)\right)^{n-2}(f_Y(y) \\ &\quad -F^{\cdot,1}(x,y))\left(f_X(x)-F^{1,\cdot}(x,y)\right)\}dxdy \end{aligned} \tag{3}$$

Proof. The proof of Corollary 1 is similar to proof of Theorem 1, in [7].

$P\{M_m(1)=l\} = P\{\text{l of the sample values in } T_2 \text{ are in } (-\infty, X_{1:n}] \times (-\infty, Y_{1:n}]\}$

Define the events E_{i_j} and $E_{i_j}^c$ as follows

$E_{i_j} = \{X_{i_j} < X_{1:n}, Y_{i_j} < Y_{1:n}\}$ and $E_{i_j}^c = \{X_{i_j} < X_{1:n}, Y_{i_j} > Y_{1:n}\} \cup \{X_{i_j} > X_{1:n}, Y_{i_j} < Y_{1:n}\} \cup \{X_{i_j} > X_{1:n}, Y_{i_j} > Y_{1:n}\}$, $1 \le i, j \le m$. Then

$$P\{M_m(1)=l\} = \sum_{i_1,i_2,\ldots,i_m} P(E_{i_1}E_{i_2}\ldots E_{i_l}E_{i_{l+1}}^c\ldots E_{i_m}^c) \tag{4}$$

By conditioning of the integral on $X=x$ and $Y=y$ in Equation (4) and using the distribution of bivariate order statistics, the proof is completed. □

When we apply the probability integral transformation $F(t)=u$, $F(s)=v$ and $F(t,s) = C_1\left(F^{-1}(t), F^{-1}(s)\right)$ and $G(t,s) = C_2\left(F^{-1}(t), F^{-1}(s)\right)$ in Equation (3), we have

$$\begin{aligned} P\{M_m(1)=l\} &= \binom{m}{l}\int_0^1\int_0^1 C_2(u,v)^l (1-C_2(u,v))^{m-l} \\ &\quad \times\{n(\widehat{C}_1(1-u,1-v))^{n-1} c_1(u,v) \\ &\quad +n(n-1)\left(\widehat{C}_1(1-u,1-v)\right)^{n-2} \\ &\quad \times(1-C_1^{\cdot}(u,v))(1-C_1^{\cdot\cdot}(u,v))\}dudv, \end{aligned} \tag{5}$$

where

$$\widehat{C}_1(1-u,1-v) = 1-u-v+C_1(u,v)$$

$$C_1^{\cdot\cdot}(u,v) = \frac{\partial C_1(u,v)}{\partial u}$$
$$C_1^{\cdot}(u,v) = \frac{\partial C_1(u,v)}{\partial v}$$
$$c_1(u,v) = \frac{\partial^2 C_1(u,v)}{\partial u \partial v}.$$

Let $C_1(u,v) = C_2(u,v)$, then

$$\begin{aligned} P\{M_m(1) = l\} \quad = \quad & \binom{m}{l} \int_0^1 \int_0^1 C_1(u,v)^l (1 - C_1(u,v))^{m-l} \\ & \times \{ n \left(\widehat{C}_1(1-u, 1-v) \right)^{n-1} c_1(u,v) \\ & + n(n-1) \left(\widehat{C}_1(1-u, 1-v) \right)^{n-2} \\ & \times (1 - C_1^{\cdot}(u,v)) (1 - C_1^{\cdot\cdot}(u,v)) \} dudv. \end{aligned} \tag{6}$$

For the ease of the calculations we firstly consider the case $C_1(u,v) = C_2(u,v) = uv$. Then we have

$$\begin{aligned} P\{M_m(1) = l\} = \binom{m}{l} \int_0^1 \int_0^1 (uv)^l (1-uv)^{m-l} \Big\{ n (1-u-v+uv)^{n-1} \\ + n(n-1)(1-u-v+uv)^{n-2} (1-u)(1-v) \Big\} dudv, \end{aligned} \tag{7}$$

In Table 1, the numerical values of $P\{M_m(1) = l\}$ are provided for $C_1(u,v) = C_2(u,v) = uv$ by using Equations (6) and (7).

Table 1. Numerical values of $P\{M_m(1) = l\}$ for different values of n and $m = 5$.

(m,n)	l	0	1	2	3	4	5
$(5,5)$	$P\{M_m(1) = l\}$	0.88	0.1	0.015	0.0021	0.00024	0.000016
$(5,10)$	$P\{M_m(1) = l\}$	0.96	0.037	0.002	0.0001	4.4×10^{-6}	1.1×10^{-6}
$(5,20)$	$P\{M_m(1) = l\}$	0.989	0.0110	0.000178	3.01×10^{-6}	4.25×10^{-8}	3.54×10^{-10}

We can interpret Table 1, as follows. For example if there have been 10 stations in a certain region in the past, after a couple of years at the same location we can observe 5 stations. So under $C_1(u,v) = C_2(u,v) = uv$, the probability of observing 1 nonhazardous station in the coming years is 0.037. In other words, the probability that only 1 station will be less than the minimum flood peak and minimum flood volume observed in the past year is 0.037, in the coming years.

In Table 2, the numerical values of $P\{M_m(1) = l\}$ are provided under $C_1(u,v) = C_2(u,v) = uv$ for $m = 10$ and some values of n. Similar to Table 1, we can do the same interpretations with Table 2. When $m = 10$ and $n = 5$, probability of observing 0 nonhazardous stations in the coming years is 0.796.

In Tables 1 and 2, it can be easily seen that while n increases, $P\{M_m(1)=l\}$ also increases for fixed values of m and l. It is clear that as l increases $P\{M_m(1)=l\}$ decreases for fixed values of m and n. Furthermore, while m increases, $P\{M_m(1)=l\}$ decreases for fixed values of n and l.

TABLE 2. Numerical values of $P\{M_m(1)=l\}$ for different values of n and $m=10$.

(m,n)	l	0	1	2	3	4	5
$(10,5)$	$P\{M_m(1)=l\}$	0.796	0.151	0.038	0.011	0.00305913	0.00085
$(10,10)$	$P\{M_m(1)=l\}$	0.926	0.0657	0.00695	0.000853	0.000109	0.0000138
$(10,20)$	$P\{M_m(1)=l\}$	0.978	0.0211	0.000739	0.0000316	1.47×10^{-6}	6.84×10^{-8}
(m,n)	l	6	7	8	9	10	
$(10,5)$	$P\{M_m(1)=l\}$	0.00022	0.000051	9.70281×10^{-6}	1.38612×10^{-6}	1.10889×10^{-7}	
$(10,10)$	$P\{M_m(1)=l\}$	1.62×10^{-6}	1.68×10^{-7}	1.43×10^{-8}	8.79×10^{-10}	2.93×10^{-11}	
$(10,20)$	$P\{M_m(1)=l\}$	3.02×10^{-9}	1.18×10^{-10}	3.81×10^{-12}	8.86×10^{-14}	1.11×10^{-15}	

The FGM copula is highly preferred in applications due to its closed form structure that facilitates theoretical calculations. In addition, it has become one of the preferred distributions in applications in the field of hydrology, since it includes both negative and positive dependency structure, see [20], [1], and [16]. For this reason, some numerical results in this paper have been calculated under the FGM copula.

Let $C_1(u,v)=uv$ and $C_2(u,v)=uv\left(1+\theta(1-u)(1-v)\right),\ \theta\in[-1,1]$, then

$$\begin{aligned} P\{M_m(1)=l\} = \binom{m}{l}\int_0^1\int_0^1 & \left[uv\left(1+\theta(1-u)(1-v)\right)\right]^l \\ & \times\left[1-uv\left(1+\theta(1-u)(1-v)\right)\right]^{m-l} \\ & \times\left\{n(1-u-v+uv)^{n-1}\right. \\ & \left. + n(n-1)(1-u-v+uv)^{n-2}(1-u)(1-v)\right\}dudv, \end{aligned} \tag{8}$$

In Table 3, the numerical values of $P\{M_m(1)=l\}$ is provided for $C_1(u,v)=uv$ and $C_2(u,v)=uv\left(1+\theta(1-u)(1-v)\right),\ \theta\in[-1,1]$ by using equation (5).

In Table 3, similar to Tables 1 and 2, as n increases $P\{M_m(1)=l\}$ also increases for fixed values of θ, m and l. For $l=0$, fixed values of m and n when the dependence parameter θ increases $P\{M_m(1)=l\}$ decreases. But for $l=1,...,5$ and fixed values of m and n, $P\{M_m(1)=l\}$ increases. As in Tables 1 and 2, while l increases, $P\{M_m(1)=l\}$ decreases for fixed values of m,n and θ.

In Table 4, the numerical values of $P\{M_m(1)=l\}$ is provided for $C_1(u,v)=C_2(u,v)=uv\left(1+\theta(1-u)(1-v)\right),\ \theta\in[-1,1]$ by using Equation (6). In Table 4, similar to Table 3 when $l=0$ and θ increases $P\{M_m(1)=0\}$ decreases for fixed values of m and n. But for $l=1,...,5$ and fixed values of m and n, $P\{M_m(1)=l\}$ increases. In Tables 5-7, the expected values of $M_m(1)$ are calculated by using Equations (5) and (6) under

$$C_1(u,v)=C_2(u,v)=uv, \tag{9}$$

TABLE 3. Numerical values of $P\{M_m(1) = l\}$ for different values of n and $m = 5$.

(m,n)	θ/l	0	1	2	3	4	5
$(5,5)$	-1	0.940	0.0533	0.00604	0.000765	0.0000836	5.61×10^{-6}
	-0.5	0.909563	0.0790261	0.00996047	0.00129879	0.000141936	9.3751×10^{-6}
	0.5	0.853735	0.122215	0.0204795	0.00316417	0.000380553	0.000026234
	1	0.828	0.140	0.0267	0.00456	0.000591	0.0000429
$(5,10)$	-1	0.988	0.0118	0.000370	0.0000146	5.31×10^{-7}	1.21×10^{-8}
	-0.5	0.974123	0.0248275	0.00100337	0.0000443511	1.68777×10^{-6}	3.89844×10^{-8}
	0.5	0.947961	0.0486292	0.00319862	0.000200988	9.89698×10^{-6}	2.75831×10^{-7}
	1	0.935	0.0595	0.00469	0.000345	0.0000195	6.10×10^{-7}
$(5,20)$	-1	0.998	0.00194	0.0000122	1.10×10^{-7}	10^{-9}	6.08×10^{-12}
	-0.5	0.993419	0.00650981	0.0000701828	8.60826×10^{-7}	9.39504×10^{-9}	6.37741×10^{-11}
	0.5	0.984326	0.0153334	0.000333565	7.29272×10^{-6}	1.28804×10^{-7}	1.30955×10^{-9}
	1	0.980	0.0196	0.000534	0.0000144	3.07×10^{-7}	3.73×10^{-9}

TABLE 4. Numerical values of $P\{M_m(1) = l\}$ for different values of n and $m = 5$.

(m,n)	θ/l	0	1	2	3	4	5
$(5,5)$	-1	0.946472	0.0487	0.00435425	0.000407848	0.0000315723	1.44141×10^{-6}
	-0.5	0.913422	0.076709	0.00876889	0.00100174	0.0000937063	5.20157×10^{-6}
	0.5	0.849052	0.124131	0.0224053	0.00384628	0.000524333	0.0000415192
	1	0.818045	0.143125	0.0312117	0.00646731	0.00105374	0.0000981558
$(5,10)$	-1	0.988971	0.0107548	0.000266524	2.01062×10^{-7}	2.01062×10^{-7}	3.14295×10^{-9}
	-0.5	0.975021	0.0240525	0.000890853	0.0000348152	1.14229×10^{-6}	2.22888×10^{-8}
	0.5	0.946488	0.04975	0.00349145	0.000258435	0.0000154613	3.1657×10^{-7}
	1	0.931939	0.061969	0.00556312	0.000492862	0.0000348765	1.40955×10^{-6}

$$C_1(u,v) = uv, C_2(u,v) = uv\left(1+\theta\left(1-u\right)\left(1-v\right)\right), \tag{10}$$

and

$$C_1(u,v) = C_2(u,v) = uv\left(1+\theta\left(1-u\right)\left(1-v\right)\right), \tag{11}$$

respectively. In Table 5, we can interpret $E\left(M_m(1)\right)$ as follows. When $m = n = 5$ (The number of stations is not changed in a certain location), expected number of nonhazardous stations is 0.139.

TABLE 5. Expected values of $M_m(1)$ for $C_1(u,v) = C_2(u,v) = uv$

m	n	$E\left(M_m(1)\right)$	m	n	$E\left(M_m(1)\right)$
5	5	0.139	5	10	0.0413
10	5	0.278	10	10	0.0826
20	5	0.556	20	10	0.165
50	5	1.39	50	10	0.413

In Table 5, we can clearly see that when m increases $E\left(M_m(1)\right)$ also increases for fixed values of n. For fixed values of m, as n increases $E\left(M_m(1)\right)$ decreases.

From Table 6, we can easily observe that as θ increases, $E(M_m(1))$ also increases for fixed values of m and n. As n increases $E(M_m(1))$ also increases for fixed values of m and θ. From Table 7, we can see that for fixed values of n and θ, when m increases $E(M_m(1))$ also increases. For fixed values of m and θ, as n increases $E(M_m(1))$ decreases. Furthermore similar to Table 7, as θ increases $E(M_m(1))$ also increases for fixed values of m and n. In Tables 8-10, the variance of $M_m(1)$

TABLE 6. Expected value of $M_m(1)$ for $C_1(u,v) = uv$, $C_2(u,v) = uv(1+\theta(1-u)(1-v))$

θ	m	n	$E(M_m(1))$	θ	m	n	$E(M_m(1))$
	5	5	0.0680		5	5	0.210
	10	5	0.0126		10	5	0.0700
-1	5	10	0.136	1	5	10	0.420
	10	10	0.0253		10	10	0.140
	20	20	0.00787		20	20	0.0828
	5	5	0.103458		5	5	0.17432
-0.5	5	10	0.206916	0.5	5	10	0.348639
	10	5	0.0269743		10	5	0.0556703
	10	10	0.0539486		10	10	0.111357

TABLE 7. Expected value of $M_m(1)$ for $C_1(u,v) = C_2(u,v) = uv(1+\theta(1-u)(1-v))$

θ	m	n	$E(M_m(1))$	θ	m	n	$E(M_m(1))$
	5	5	0.0587987		5	5	0.229656
	10	5	0.117597		10	5	0.459311
-1	5	10	0.0113121	1	5	10	0.0747203
	5	5	0.0976528		5	5	0.182785
-0.5	10	5	0.195306	0.5	10	5	0.365559
	5	10	0.0259433		5	10	0.0575716

are calculated by using Equations (5) and (6) under

$$C_1(u,v) = C_2(u,v) = uv, \tag{12}$$

$$C_1(u,v) = uv, C_2(u,v) = uv(1+\theta(1-u)(1-v)), \tag{13}$$

and

$$C_1(u,v) = C_2(u,v) = uv(1+\theta(1-u)(1-v)), \tag{14}$$

respectively.

In Table 8, it is obvious that for fixed values of n, when m increases variance of $M_m(1)$ also increases. For fixed values of m, as n increases the variance of $M_m(1)$

decreases. Similary, in Tables 9 and 10, for fixed values of θ and n, as m increases variance of $M_m(1)$ increases. Furthermore for fixed values of m and n, when θ increases, the variance of $M_m(1)$ increases.

TABLE 8. Variances of $M_m(1)$ for $C_1(u,v) = C_2(u,v) = uv$

m	n	$V(M_m(1))$	m	n	$V(M_m(1))$
5	5	0.165	5	10	0.044
10	5	0.404	10	10	0.097
20	5	1.1	20	10	0.226

TABLE 9. Variances of $M_m(1)$ for $C_1(u,v) = uv$, $C_2(u,v) = uv\,(1+\theta\,(1-u)\,(1-v))$

θ	m	n	$V\,(M_m(1))$	θ	m	n	$V\,(M_m(1))$
	5	5	0.0812		5	5	0.254
	10	5	0.198		10	5	0.642
-1	5	10	0.133	1	5	10	0.0768
	10	10	0.0284		10	10	0.173
	5	5	0.121		5	5	0.201
-0.5	5	10	0.0286	0.5	5	10	0.0603
	10	5	0.297		10	5	0.52
	10	10	0.061		10	10	0.134

TABLE 10. Variances of $M_m(1)$ for $C_1(u,v) = C_2(u,v) = uv\,(1+\theta\,(1-u)\,(1-v))$

θ	m	n	$V\,(M_m(1))$	θ	m	n	$V\,(M_m(1))$
	5	5	0.0669047		5	5	0.292749
	5	10	0.0117666		5	10	0.0836673
-1	10	5	0.1558	1	10	5	0.769603
	5	5	0.112893		5	5	0.224385
-0.5	5	10	0.027275	0.5	5	10	0.0629825
	10	5	0.268656		10	5	0.5694260

3. CONCLUSION

In this study, a bivariate exceedance model is constructed based on bivariate order statistics. In this model, we compose a bivariate random threshold model by using the past flood peak and flood volume of the hydrological stations. Probability of exceedance statistics are calculated under some well-known copulas for small sample sizes. Then the numerical values of expected values of exceedance statistics are provided for independent and FGM copulas. Because of the complexity of the calculations, the numerical results are provided for small sample sizes. As a further study, we need to investigate the properties of exceedance statistics under different bivariate distributions by using some real data sets in hydorology. The results obtained using real data sets can be compared with the theoretical results in this article.

Declaration of Competing Interests The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this manuscript.

Acknowledgements The author thanks to the referees and editors for their valuable comments which resulted in the improvement of the paper.

References

[1] Ariff, N. M., Jemain, A. A., Ibrahim, K., Zin, W. W., IDF relationships using bivariate copula for storm events in Peninsular Malaysia. *Journal of Hydrology*, 470 (2012), 158-171. https://doi.org/10.1016/j.jhydrol.2012.08.045.

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

[3] Barakat, H. M., On moments of bivariate order statistics. *Annals of the Institute of Statistical Mathematics*, 51(2) (1999), 351-358.

[4] Bayramoglu, I., Eryilmaz, S., Order statistics of dependent sequences consisting of two different sets of exchangeable variables. *Journal of Computational and Applied Mathematics*, 286 (2015), 1-6. https://doi.org/10.1016/j.cam.2015.02.045.

[5] Bayramoglu, I.,Giner, G., Order statistics and exceedances for some models of INID random variables, *Brazilian Journal of Probability and Statistics*, 28 (4) (2014), 492-514. https://doi.org/10.1214/13-BJPS221.

[6] Bairamov, I. G., Petunin, Y. I., Statistical tests based on training samples, *Cybernetics and Systems Analysis,* 27 (3) (1991), 408-413.

[7] Erem, A., Bayramoglu, I., Exact and asymptotic distributions of exceedance statistics for bivariate random sequences. *Statistics & Probability Letters,* 125 (2017), 181-188. https://doi.org/10.1016/j.spl.2017.02.012.

[8] Erem, A., Bivariate two sample test based on exceedance statistics. *Communications in Statistics-Simulation and Computation, 49(9)* (2020), 2389-2401. https://doi.org/10.1016/j.spl.2017.02.012

[9] Erem, A., Bayramoglu, I., Bivariate general random threshold models and exceedance statistics. *TWMS Journal of Pure and Applied Mathematics*, 11(2) (2020), 189-203.

[10] Eryılmaz, S., Bairamov, I. G., On a new sample rank of an order statistics and its concomitant. *Statistics & Probability Letters*, 63(2) (2003), 123-131. https://doi.org/10.1016/S0167-7152(03)00059-2

[11] Eryilmaz, S., The longest run statistic associated with exchangeable binary variables. *Turkish Journal of Engineering and Environmental Sciences*, 29(2) (2005), 105-112.

[12] Eryilmaz, S., Gebizlioglu, O. L., Tank, F., Modeling of claim exceedances over random thresholds for related insurance portfolios. *Insurance: Mathematics and Economics*, 49(3) (2011), 496-500. https://doi.org/10.1016/j.insmatheco.2011.08.009

[13] Huang, Y., Liang, Z., Hu, Y., Li, B., Wang, J., Theoretical derivation for the exceedance probability of corresponding flood volume of the equivalent frequency regional composition method in hydrology. *Hydrology Research*, 51(6) (2020), 1274-1292. https://doi.org/10.2166/nh.2020.027

[14] Kemalbay, G., Bayramoglu, I., Joint distribution of new sample rank of bivariate order statistics. *Journal of Applied Statistics*, 42(10) (2015), 2280-2289. https://doi.org/10.1080/02664763.2015.1023705

[15] Nathan, R., Jordan, P., Scorah, M., Lang, S., Kuczera, G., Schaefer, M., Weinmann, E. Estimating the exceedance probability of extreme rainfalls up to the probable maximum precipitation, *Journal of hydrology*, 543 (2016), 706-720. https://doi.org/10.1016/j.jhydrol.2016.10.044.

[16] Papaioannou, G., Kohnová, S., Bacigál, T., Szolgay, J., Hlavčová, K., Loukas, A., Joint modelling of flood peaks and volumes: A copula application for the Danube River. *Journal of Hydrology and Hydromechanics*, 64(4) (2016), 382-392. https://doi.org/10.1515/johh-2016-0049

[17] Stoimenova, E., The power of exceedance-type tests under lehmann alternatives. *Communications in Statistics-Theory and Methods*, 40(4) (2011), 731-744. https://doi.org/10.1080/03610920903453475

[18] Stoimenova, E., Balakrishnan, N., A class of exceedance-type statistics for the two-sample problem. *Journal of Statistical Planning and Inference*, 141 (9) (2011), 3244-3255. https://doi.org/10.1080/03610920903453475

[19] Thompson, J. A., Bissett, W. T., Sweeney, A. M., Evaluating geostatistical modeling of exceedance probability as the first step in disease cluster investigations: very low birth weights near toxic Texas sites. *Environmental Health*, 13(1) (2014), 1-6.

[20] Yue, S., Ouarda, T. B., Bobée, B., A review of bivariate gamma distributions for hydrological application. *Journal of Hydrology*, 246(1-4) (2001), 1-18. https://doi.org/10.1016/S0022-1694(01)00374-2

[21] Wesolowski, J., Ahsanullah, M., Distributional properties of exceedance statistics. *Ann. Inst. Statist. Math.*, 50 (1998), 543-565.