BIVARIATE RANDOM SEQUENCES AND EXACT AND ASYMPTOTIC DISTRIBUTIONS OF EXCEEDANCE STATISTICS

AYŞEGÜL EREM

JUNE 2017

BIVARIATE RANDOM SEQUENCES AND EXACT AND ASYMPTOTIC DISTRIBUTIONS OF EXCEEDANCE STATISTICS

A DISSERTATION SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL

AND APPLIED SCIENCES OF

IZMIR UNIVERSITY OF ECONOMICS

BY

AYŞEGÜL EREM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

JUNE 2017

Ph.D. DISSERTATION EXAMINATION RESULT FORM

Approval of the Graduate School of Natural and Applied Sciences

Assoc. Prof. Dr. Devrim Ünay
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Gözde Yazgı Tütüncü
Head of Department

We have read the dissertation entitled **"Bivariate random sequences and exact and asymptotic distributions of exceedance statistics"** completed by **AYŞEGÜL EREM** under supervision of **Prof. Dr. İsmihan Bayramoğlu** and we certify that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Dr. İsmihan Bayramoğlu
Supervisor

Examining Committee Members **Date:** 06.06.2017

Prof. Dr. İsmihan Bayramoğlu
Dept. of Mathematics, IUE

Prof. Dr. Yılmaz Akdi
Dept. of Statistics, Ankara University

Prof. Dr. Özlem Ege Oruç
Dept. of Statistics, DEU

Prof. Dr. Gözde Yazgı Tütüncü
Dept. of Mathematics, IUE

Assoc. Prof. Dr. Femin Yalçın
Dept. of Engineering Science, IKCU

ABSTRACT

BIVARIATE RANDOM SEQUENCES AND EXACT AND ASYMPTOTIC DISTRIBUTIONS OF EXCEEDANCE STATISTICS

AYŞEGÜL EREM
Ph.D. in Applied Mathematics and Statistics
Graduate School of Natural and Applied Sciences
Supervisor: Prof. Dr. İsmihan Bayramoğlu
June 2017

In this thesis, random threshold models based on bivariate random sequences are investigated. The random threshold models are defined on the basis of rth order statistic and rth concomitant. The finite and asymptotic distributions of exceedance statistics are obtained. Distributions of exceedance statistics presented in this work, do not depend on marginal distribution functions and depend only on copulas. Applications of these bivariate random threshold models in medicine and air pollution are discussed.

Keywords: Random threshol models, exceedance statistics, order statistics, concomitants, copula, asymptotic distribution.

ÖZ

İKİ DEĞİŞKENLİ RASGELE DİZİLER VE AŞAN İSTATİSTİKLERİNİN SONLU VE ASİMPTOTİK DAĞILIMLARI

AYŞEGÜL EREM
Uygulamalı Matematik ve İstatistik, Doktora
Fen Bilimleri Enstitüsü
Tez Danışmanı: Prof. Dr. İsmihan Bayramoğlu
Haziran 2017

Bu tezde, iki değişkenli rasgele değişken dizilerine dayalı rasgele bariyer modelleri incelenmiştir. Rasgele bariyer modelleri r' ıncı sıra istatistiğine ve eşleniğine dayalı olarak tanımlanmıştır. Aşan istatistiklere ilişkin sonlu ve asimptotik dağılımlar elde edilmiştir. Bu çalışmada sunulan aşan istatistiklerin dağılımları marjinal dağılımlara bağlı değildir ve yalnızca kapulalara bağlıdır. Bu iki değişkenli rasgele bariyer modellerinin sağlık ve hava kirliliğindeki uygulamaları ortaya konulmuştur.

Anahtar Kelimeler: Rasgele bariyer modelleri, aşan istatistikler, sıra istatistikleri, eşlenikler, kapula, asimptotik dağılım.

ACKNOWLEDGEMENT

First and foremost, I would like to extend my heartfelt gratitude to my PhD advisor, Prof. Dr. İsmihan Bayramoğlu for his immense academic knowledge, encouragement and great contributions during my PhD process. It has been a great honour and a privilege to be his student.

I would also like to express my special thanks to Prof. Dr. Gözde Yazgı Tütüncü and Asst. Prof. Dr. Aslı Güldürdek for their timeless support and guidance in every aspect of my life.

I sincerely thank my colleagues in Izmir University of Economics for their great friendship, and for many great memories.

I deeply thank to my family for their unconditional support throughout my life. I couldn' t have been able to finish this journey without their constant encouragement and support.

Finally, I would like to thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for supporting me with its national PhD scholarship programme.

To my family and in memory of my beloved grandmother ...

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Random threshold models have aroused much scientific interest, due to the wide application areas. These models can be used in many areas, such as medicine, engineering, modelling environment events (flooding see [36], air and water pollution, etc.), and insurance. For some interesting applications of exceedance statistics in insurance, see [37] and [27].

Additionally, these models are used in quality control to detect extreme variation in some quality characteristics, see e.g. [69], [70], and [57].

Since 1940s, in the statistical literature, there have appeared numerous paper on exceedance statistics and random threshold models (see e.g. [41], [39], [40], [34], [8], [60], and [58]).

Exceedance statistics play a fundamental role in nonparametric methods in construction of distribution free tests, see [18], [59], [45], [46], [43], [44], [52], and [53]. In some exceedance statistics studies, based on ordered statistics, a new efficient test criterion was constructed for testing hypothesis H_0 against several class of alternatives. Katzenbeisser [45], [46] constructed a test based on exceedance statistics for two sample problem in univariate independent samples. In his studies, the test statistic based on exceedances were used for testing homogeneity of two independent random samples. This new test in the case of the stochastically ordered alternatives appeared to be unbiased and efficient.

Eryilmaz et al. [37] used exceedance statistics in modeling claim exceedances over random thresholds for insurance portfolios. Eryilmaz et al. [37] considered two different models of portfolio claims: independent and identically distributed portfolio claims, and exchangeable dependent portfolio claims.

We also refer to Benestad [23], [24], [25] for application of exceedance statistics in hydrological and climate events.

Bairamov [9] considered univariate random threshold model based on independent and identically distributed (iid) random variables and upper record values. In his paper, exact and asymptotic distributions of exceedances statistics were given and some distribution free properties of exceedance statistics were presented.

Wesolowski et al. [66] considered a random variable with a distribution function (cdf) F and a sequences of iid random variables with the cdf G. For three different exceedance statistics defined on the base of order statistics and record statistics, Wesolowski et al. [66] provide exact and asymptotic distributions. This study proves some distribution free properties of considered exceedance statistics and also provides a characterization of equidistribution of two independent samples.

Bairamov et al. [14] defined new exceedance models based on generalized order statistics and discussed applications of these models in life time data analysis for different censoring schemes. Bairamov et al. [15], investigated the joint behavior of exceedance and precedance statistics. Bairamov et al. [16] considered the waiting time of exceedances in a random threshold model based on ordered statistics.

Turhan [62] considered two independent random samples and, used the rth and sth ($r < s$) order statistics to introduce new exceedance statistics. The exact and asymptotic distributions of these statistics were used for construction of nonparametric test criterions for testing homogeneity of two independent random samples.

Bayramoglu and Giner [22] introduced new threshold models based on independent but not necessarily identically distributed random variables. They studied the finite and asymptotic distributions of exceedance statistics and provided different numerical results, as well as a discussion of possible practical applications in insurance.

Although there have been many studies on univariate random threshold models, few studies consider multivariate random threshold models. In this thesis, we focus on random exceedance models based on bivariate samples. In particular, we investigate the exact and asymptotic distributions of exceedance statistics constructed for bivariate random samples based on order statistics and concomitants of these samples. The properties and applications of these exceedance statistics are also the subject of this thesis.

This thesis is organised as follows. Definitions of order statistics, concomitants and record statistics are given in Chapter 2. The distribution theory of ordered statistics are also described for both independent and identically distributed (iid) and independent but not necessarily identically distributed (inid) random variables. Also, some distributional properties of records and concomitants for iid random variables are given.

Chapter 3 consists of a comprehensive literature review of exceedance statistics. Firstly, the random threshold models from iid random variables are considered. After this, we consider the random threshold models based on order statistics, concomitants and records, providing distribution free properties of random thresholds constructed on records. Applications of these exceedance statistics based on invariant confidence interval and the random threshold based on inid random variables are discussed. At the end of this chapter, some exceedance statistics from dependent random variables are presented.

In Chapter 4, new random threshold models are introduced based on bivariate random variables. After the introduction of a new random threshold model constructed on random variables (X, Y), the finite and asymptotic distributions of exceedance statistics are derived, as well as distributions of normalized exceedance statistics. Next, new random threshold models based on rth order statistics and rth concomitants are introduced. The finite and asymptotic distributions of new exceedance statistics are derived and distribution of normalized exceedance statistics are obtained.

In Chapter 5, we discuss ways in which with new random threshold models introduced in Chapter 4 can be applied in medicine and air pollution.

Chapter 2

Ordered Random Variables

In this chapter, we will give brief definitions of order statistics of independent and identically distributed, independent but not necesserily identically distributed random variables, concomitants, record statistics as well as some useful formulas based on these statistics.

2.1 Order statistics

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space. In this probability space, we consider random variables $X_i(w) \equiv X_i$, $w \in \Omega$ and we assume that these random variables are independent copies of a random variable $X(w)$, $w \in \Omega$. We assume that other random variables considered in this thesis are also defined in same probability space $\{\Omega, \mathcal{F}, P\}$.

2.1.1 Order Statistics of iid random variables

For many years, order statistics attracted the interest of many statisticians. Due to their extensive applications in many areas, such as hypothesis testing, estimation of parameters, reliability engineering, survival analysis, biology and medicine, finance, and economics. The order statistics have been one of the essential topics of probability and statistics in the last semicentennial. We refer to David and Nagaraja [32], [33] and Arnold et al. [1] for a detailed review on order statistics.

Let $X_1, X_2, ..., X_n$ be iid random variables with cdf $F_X(x)$. If we order those random variables in ascending order, we have the order statistics based on the sample $X_1, X_2, ..., X_n$ as follows

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n},$$

where $X_{1:n}$ denotes the smallest order statistics, $X_{r:n}$, $1 \leq r \leq n$, denotes the rth order statistic, and $X_{n:n}$ denotes the largest order statistic.

If X_i' s are independent and identically distributed random variables, then the cumulative distribution function (cdf) of rth order statistic is given by (see Arnold et al., 2008; p. 12)

$$F_{X_{r:n}}(x) = \sum_{i=r}^{n} \binom{n}{i} \left[F_X(x)\right]^i \left[1 - F_X(x)\right]^{n-i}. \tag{2.1}$$

It is known that equation (2.1) holds for both discrete and continuous random variables. If X is continuous random variable with cdf $F_X(x)$ and probability density function (pdf) $f_X(x)$ then the pdf of rth order statistic is

$$f_{X_{r:n}}(x) = \frac{1}{Beta(r, n-r+1)} \left[F_X(x)\right]^{r-1} \left[1 - F_X(x)\right]^{n-r} f_X(x),\ 1 \leq r \leq n, \tag{2.2}$$

where $Beta(r, n-r+1)$ is a beta function (see Casella and Berger, 2002; p. 229).

In absolutely continuous case, for $1 \leq r < s \leq n$ joint density function of rth and sth order statistics is given by (see David and Nagaraja, 2003; p. 12)

$$f_{X_{r:n},X_{s:n}}(x,y)=\begin{cases}\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left[F_X(x)\right]^{r-1}f_X(x) & \\ \times[F_Y(y)-F_X(x)]^{s-r-1}f_Y(y)\left[1-F_Y(y)\right]^{n-s}, & x\leq y\\ 0, & \text{otherwise}\end{cases}$$

Furthermore, for $x<y$ the joint distribution function of rth and sth order statistics is

$$F_{X_{r:n},X_{s:n}}(x,y)=\begin{cases}\sum_{j=s}^{n}\sum_{i=r}^{j}\frac{n!}{i!(j-i)!(n-j)!}\left[F_X(x)\right]^{i} & \\ \times[F_Y(y)-F_X(x)]^{j-i}\left[1-F_Y(y)\right]^{n-j}, & x<y\\ F_{X_{s:n}}(y), & x\geq y\end{cases}\tag{2.3}$$

It is known that equation (2.3) holds for both continuous and discrete random variables.

Also, for $1\leq r_1<r_2<\cdots<r_k\leq n$, the joint pdf of first k order statistics, $X_{r_1:n},X_{r_2:n},...,X_{r_k:n}$ is

$$f_{X_{r_1:n},...,X_{r_k:n}}(x_1,x_2,...,x_k)\tag{2.4}$$
$$=\begin{cases}\frac{n!}{(r_1-1)!(r_2-r_1-1)!...(n-r_k)!}\left[F_X(x_1)\right]^{r_1-1} & \\ \times\left[F_X(x_2)-F_X(x_1)\right]^{r_2-r_1-1}\left[F_X(x_3)-F_X(x_2)\right]^{r_3-r_2-1} & \text{if } x_1\leq x_2\leq ...\leq x_k\\ \times\cdots\left[1-F_X(x_k)\right]^{n-r_k}f_X(x_1)\cdots f_X(x_k), & \\ 0, & \text{otherwise}\end{cases}$$

The joint pdf of all n order statistics is

$$f_{X_{1:n},...,X_{n:n}}(x_1,...,x_n)=\begin{cases}n!f(x_1)\cdots f(x_n), & x_1\leq\cdots\leq x_n\\ 0, & \text{otherwise}\end{cases}\tag{2.5}$$

If X_i' s are discrete random variables, then the probability mass function (pmf) of the rth ordered statistics is (see David and Nagaraja, 2003; p. 16)

$$\begin{aligned} P\{X_{r:n}=x\} &= F_{X_{r:n}}(x)-F_{X_{r:n}}(x-1) \\ &= \frac{1}{B(r,n-r+1)}\left(I_{F_X(x)}(r,n-r+1)-I_{F_X(x-1)}(r,n-r+1)\right) \\ &= \frac{1}{B(r,n-r+1)}\int_{F_X(x-1)}^{F_X(x)} z^{r-1}(1-z)^{n-r}dz, \end{aligned} \tag{2.6}$$

where $I_{F_{X_{r:n}}(x)}(r,n-r+1)$ is an incomplete beta function given by

$$I_{F_{X_{r:n}}(x)}(r,n-r+1)=\int_0^{F_X(x)} z^{r-1}(1-z)^{n-r}dz.$$

Therefore,

$$P\{X_{r:n}=x\}=\begin{cases} \frac{1}{B(r,n-r+1)}\int_{F_X(x-1)}^{F_X(x)} z^{r-1}(1-z)^{n-r}dz, & x=0,1,2,... \\ 0, & \text{otherwise} \end{cases}$$

Furthermore, the pmf of the rth and sth order statistics is

$$f_{X_{r:n},X_{s:n}}(x,y)=\begin{cases} F_{X_{r:n},X_{s:n}}(x,y)-F_{X_{r:n},X_{s:n}}(x-1,y) & x\le y \\ -F_{X_{r:n},X_{s:n}}(x,y-1)+F_{X_{r:n},X_{s:n}}(x-1,y-1), & \\ 0, & \text{otherwise} \end{cases}$$

2.1.2 Order statistics of inid random variables

Let $X_1, X_2, ..., X_n$ be a sequence of independent random variables with continuous cdf's $F_1(x), F_2(x), ..., F_n(x)$, respectively. Assume that $f_i(x) = \frac{dF_i(x)}{dx}$, $i = 1, ..., n$. Then for $1 \leq r \leq n$, the probability density function of rth order statistic $X_{r:n}$ is (see David and Nagaraja, 2003; p. 96)

$$f_{X_{r:n}}(x) = C_{n,r} \sum_{i_1, i_2, \ldots, i_n} \prod_{l=1}^{r-1} f_{i_l}(x) f_{i_r}(x) \prod_{j=r+1}^{n} \left(1 - F_{i_j}(x)\right) \quad -\infty < x < \infty, \tag{2.7}$$

where $\sum\limits_{i_1, i_2, \ldots, i_n}$ denotes the sum over all $n!$ permutations $(i_1, i_2, ..., i_n)$ of n and $C_{n,r} = \frac{1}{(r-1)!(n-r)!}$.

For $1 \leq k_1 < \cdots < k_n \leq n$ and $1 \leq r < s \leq n$ the joint pdf of $X_{r:n}$ and $X_{s:n}$ is

$$f_{X_{r:n}, X_{s:n}}(x_1, x_2) = \begin{cases} C_{n,r,s} \sum\limits_{k_1, k_2, \ldots, k_n} F_{k_1}(x_1) \cdots F_{k_{r-1}}(x_1) f_{k_r}(x_1) & \\ \quad \times \left\{F_{k_{r+1}}(x_2) - F_{k_{r+1}}(x_1)\right\} \cdots & x_1 \leq x_2 \\ \quad \times \left\{F_{k_{n-1}}(x_2) - F_{k_{n-1}}(x_1)\right\} f_{k_s}(x_2) & \\ \quad \times \left\{1 - F_{k_{s+1}}(x_2)\right\} \cdots \left\{1 - F_{k_n}(x_2)\right\}, & \\ 0, & \text{otherwise} \end{cases} \tag{2.8}$$

where $\sum\limits_{k_1, k_2, \ldots, k_n}$ denotes the sum over all $n!$ permutations $(k_1, k_2, ..., k_n)$ of n and

$C_{n,r,s} = \frac{1}{(r-1)!(s-r-1)!(n-s)!}$ (for more detail see the work of Balakrishnan [19]).

As it can be seen from the equations (2.7) and (2.8), for inid random variables density functions of order statistics are quite complicated. Therefore, Vaughan and Venables [64] expressed the distributions of order statistics in terms of permanents.

Definition 2.1.1 (Permanents)**.** (See Aitken, 1939; p. 30) Consider a square matrix of order $n \times n$, such that

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Then the permanent of matrix A is

$$perm\,(A) = \sum_{\wp_{1,2,\ldots,m}} \prod_{j=1}^{m} a_{ji_j}, \tag{2.9}$$

where $\sum_{\wp_{1,2,\ldots,m}}$ is the class of sum over all $m!$ permutations $(i_1, i_2, ..., i_n)$ of $(1, 2, ..., m)$.

The permanent of a square matrix is similar to the determinant with difference of all signs are positive. Also, it is usually denoted by $^+\,|A|^+$.

For $1 \leq r \leq n$, the marginal density function of $X_{r:n}$ is (see Vaughan and Venables [64])

$$f_{X_{r:n}}(x) = \frac{1}{(r-1)!\,(n-r)!} Perm\,(A_1)\,, \quad -\infty < x < \infty,$$

where

$$A_1 = \begin{bmatrix} F_1(x) & \cdots & F_n(x) \\ \vdots & & \vdots \\ F_1(x) & \cdots & F_n(x) \\ f_1(x) & \cdots & f_n(x) \\ 1-F_1(x) & \cdots & 1-F_n(x) \\ \vdots & & \vdots \\ 1-F_1(x) & \cdots & 1-F_n(x) \end{bmatrix} \begin{array}{l} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \begin{matrix} r-1 \\ \text{rows} \end{matrix} \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \begin{matrix} n-r \\ \text{rows} \end{matrix} \end{array} .$$

For $1 \leq r < s \leq n$, the joint pdf of rth and sth order statistics is (see Vaughan and Venables [64])

$$f_{X_{r:n},X_{s:n}}(x_1, x_2) = \begin{cases} \frac{1}{(r-1)!(s-r-1)!(n-s)!} Perm(A_2), & x_1 \leq x_2 \\ 0, & \text{otherwise} \end{cases}$$

where

$$A_2 = \begin{bmatrix} F_1(x_1) & F_2(x_1) & \cdots & F_n(x_1) \\ \vdots & \vdots & & \vdots \\ F_1(x_1) & F_2(x_1) & \cdots & F_n(x_1) \\ f_1(x_1) & f_2(x_1) & \cdots & f_n(x_1) \\ F_1(x_2)-F_1(x_1) & F_2(x_2)-F_2(x_1) & \cdots & F_n(x_2)-F_n(x_1) \\ \vdots & \vdots & & \vdots \\ F_1(x_2)-F_1(x_1) & F_2(x_2)-F_2(x_1) & \cdots & F_n(x_2)-F_n(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_n(x_2) \\ 1-F_1(x_2) & 1-F_2(x_2) & \cdots & 1-F_n(x_2) \\ \vdots & \vdots & & \vdots \\ 1-F_1(x_2) & 1-F_2(x_2) & \cdots & 1-F_n(x_2) \end{bmatrix} \begin{array}{l} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \begin{matrix} r-1 \\ \text{rows} \end{matrix} \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \begin{matrix} s-1 \\ \text{rows} \end{matrix} \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \begin{matrix} n-s \\ \text{rows} \end{matrix} \end{array}$$

For $1 \leq k_1 < k_2 < \cdots < k_n \leq n$, joint density function of any subset

$X_{k_1:n}, X_{k_2:n}, \cdots, X_{k_r:n}$ is

$$f_{X_{k_1:n}, X_{k_2:n}, \ldots, X_{k_r:n}}(x_1, x_2, \ldots, x_r) = \begin{cases} C_{k_1,k_2,\ldots,k_r} Perm\left(A_3\right), & x_1 \leq x_2 \leq \cdots \leq x_n \\ 0, & \text{otherwise} \end{cases},$$

where

$$C_{k_1,k_2,\ldots,k_r} = \frac{1}{(k_1-1)!(k_2-k_1-1)!\cdots(k_r-k_{r-1}-1)!\,(n-k_r)!},$$

and

$$A_3 = \left[\begin{array}{ccc} F_1(x_1) & \cdots & F_n(x_n) \\ \vdots & & \vdots \\ F_1(x_1) & \cdots & F_n(x_n) \\ f_1(x_1) & \cdots & f_n(x_1) \\ F_1(x_2)-F_1(x_1) & \cdots & F_n(x_2)-F_n(x_1) \\ \vdots & & \vdots \\ F_1(x_2)-F_1(x_1) & \cdots & F_n(x_2)-F_n(x_1) \\ f_1(x_2) & \cdots & f_n(x_2) \\ \vdots & & \vdots \\ F_1(x_k)-F_1(x_{k-1}) & \cdots & F_n(x_k)-F_n(x_{k-1}) \\ \vdots & & \vdots \\ F_1(x_k)-F_1(x_{k-1}) & \cdots & F_n(x_k)-F_n(x_{k-1}) \\ f_1(x_k) & \cdots & f_n(x_k) \\ 1-F_1(x_k) & \cdots & 1-F_n(x_k) \\ \vdots & & \vdots \\ 1-F_1(x_k) & \cdots & 1-F_n(x_k) \end{array}\right] \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \begin{array}{c} k_1-1 \\ \text{rows} \end{array} \\ \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \begin{array}{c} k_2-k_1-1 \\ \text{rows} \end{array} \\ \\ \qquad\quad\vdots \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \begin{array}{c} k_r-k_{r-1}-1 \\ \text{rows} \end{array} \\ \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \begin{array}{c} n-k_r-1 \\ \text{rows} \end{array} \end{array}$$

Furthermore, joint density function of $X_{1:n}, X_{2:n}, ..., X_{n:n}$ is

$$f_{X_{1:n},X_{2:n},...,X_{n:n}}(x_1, x_2, ..., x_n) = \begin{cases} Perm(A_4), & x_1 \leq x_2 \leq \cdots \leq x_n \\ 0, & \text{otherwise} \end{cases},$$

where

$$A_4 = \begin{bmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & & \vdots \\ f_1(x_n) & \cdots & f_n(x_n) \end{bmatrix}.$$

2.2 Concomitants of order statistics

In pursuit of the improvement in theory of order statistics, concomitants were introduced related with ordered random variables. Concomitants have been widely used in statistical inference based on bivariate random samples.

Let $Z_i = (X_i, Y_i)$, $i = 1, 2, ..., n$, be an absolutely continuous bivariate random sequence with joint cdf $F(x, y)$ and joint pdf $f(x, y)$. Let $F_X(x)$ and $F_Y(y)$ be the marginal cdf's of X_i and Y_i, respectively. Let $f_X(x) = \frac{dF_X(x)}{dx}$ and $f_Y(y) = \frac{dF_Y(y)}{dy}$, be the corresponding pdf's of X and Y. If the sample Z_i is ordered by X_i, then the corresponding pair of $Y-$variate order statistics is called rth concomitants of rth order statistics and denoted by $X_{[r:n]}$. The theory of concomitants is well documented in David et al. [31], David and Nagaraja [32], [33], and He and Nagaraja [42].

Let $(X_{r:n}, Y_{[r:n]})$ be a vector of rth order statistic and its concomitant constructed from the sequence Z_i. The joint pdf of $X_{r:n}$ and $Y_{[r:n]}$ is given by

$$f_{X_{r:n},Y_{[r:n]}}(x,y) = f(y \mid x) f_{X_{r:n}}(x), \tag{2.10}$$

where $f_{X_{r:n}}(x)$ is the pdf of rth order statistic, X_i and $f(y \mid x)$ is the conditional pdf of Y given $X = x$.

The marginal density function of the rth concomitant of rth order statistic can be easily obtained from equation (2.10) as follows:

$$f_{Y_{[r:n]}}(y) = \int_{-\infty}^{\infty} f(y \mid x) f_{X_{r:n}}(x) dx. \tag{2.11}$$

For $r < s$ the joint pdf of the rth and sth concomitants is

$$f_{Y_{[r:n]},Y_{[s:n]}}(y_1, y_2) = \int_{-\infty}^{\infty} \int_{x_1}^{\infty} f(y_2 \mid x_2) f(y_1 \mid x_1) f_{X_{r:n},X_{s:n}}(x_1, x_2) dx_2 dx_1, \tag{2.12}$$

where $f_{X_{r:n},X_{s:n}}(x_1, x_2)$ is the joint pdf of $X_{r:n}$ and $X_{s:n}$. More generally, for $1 \leq r_1 < \cdots < r_j \leq n$, the joint pdf of $Y_{[r_1:n]}, ..., Y_{[r_n:n]}$ is

$$\begin{aligned} &f_{Y_{[r_1:n]},...,Y_{[r_n:n]}}(y_1, ..., y_k) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{x_{r_k}} \cdots \int_{-\infty}^{x_{r_2}} \prod_{i=1}^{k} f(y_{r_i} \mid x_{r_i}) f_{X_{r_1:n},...,X_{r_k:n}}(x_{r_1}, ..., x_{r_k}) dx_{r_1} \cdots dx_{r_k}, \end{aligned} \tag{2.13}$$

where $f_{X_{r_1:n},...,X_{r_k:n}}(x_1, ..., x_k)$ is the joint pdf of $X_{r_1:n}, ..., X_{r_k:n}$.

From equations (2.12) and (2.13), one can easily calculate the expected value and variance of the rth concomitant as (see Yang [65])

$$E(Y_{[r:n]}) = E(E(Y_1 \mid X_1 = X_{r:n})), \tag{2.14}$$

$$Var(Y_{[r:n]}) = E(Var(Y_1 \mid X_1 = X_{r:n})) + Var(E(Y_1 \mid X_1 = X_{r:n})). \tag{2.15}$$

Here

$$E\left(E\left(Y_1 \mid X_1 = X_{r:n}\right)\right) = E\left(\psi\left(X_{r:n}\right)\right),$$

where $\psi(X) = E(Y_1 \mid X_1 = x)$. Likewise,

$$E(Var(Y_1 \mid X_1 = X_{r:n})) = E(\psi(X_{r:n})),$$

where $\psi(X) = Var(Y_1 \mid X_1 = x)$.

Similarly, covariance of $Y_{[r:n]}$ and $Y_{[s:n]}$ and the covariance of rth order statistic and sth concomitant is

$$Cov(Y_{[r:n]}, Y_{[s:n]}) = Cov\left[E(Y_1 \mid X_1 = X_{r:n}), E(Y_1 \mid X_1 = X_{s:n})\right] \ (r \neq s),$$
$$Cov(X_{r:n}, Y_{[s:n]}) = Cov\left[X_{r:n}, E(Y_1 \mid X_1 = X_{s:n})\right].$$

Definition 2.2.1 (Positive Quadrant Dependence)**.** (See Lehmann [50]) Let (X, Y) be a pair of random variable with cdf $G(x, y)$ and marginals $F_X(x)$ and $F_Y(y)$, respectively. Then for all x and y, the random variables (X, Y) are called positive quadrant dependent, if

$$G(x, y) \geq F_X(x) F_Y(y), \ \forall \ (x, y) \in \mathbb{R}^2. \tag{2.16}$$

It is obvious that equation (2.16) can be also written in the following forms:

$$P(X \leq x, Y \geq y) \geq P(X \leq x) P(Y \geq y), \forall \ (x, y) \in \mathbb{R}^2 \tag{2.17}$$
$$P(X \geq x, Y \leq y) \geq P(X \geq x) P(Y \leq y), \forall \ (x, y) \in \mathbb{R}^2 \tag{2.18}$$
$$P(X \geq x, Y \geq y) \geq P(X \geq x) P(Y \geq y), \forall \ (x, y) \in \mathbb{R}^2. \tag{2.19}$$

Let $X_1, X_2, ..., X_n$ be an iid random sample. If $E(Y \mid X = x)$ is monotone, then for any x_1, x_2 and $r, s = 1, 2, ..., n$, $\left(X_{[r:n]}, X_{[s:n]}\right)$ are positive quadrant dependent (see Kim and David [48]).

$$P\left\{X_{[r:n]} \geq x_1, X_{[s:n]} \geq x_2\right\} \geq P\left\{X_{[r:n]} \geq x_1\right\} P\left\{X_{[s:n]} \geq x_2\right\}, \forall \ (x, y) \in \mathbb{R}^2.$$

2.3 Record statistics

Record statistics have been aroused great interest among statisticians. Further, in recent years, it is widely used in reliability theory, risk analysis, and modelling events in insurance, flooding, geosciences, sports, and climate science (see Benestad [25], [24] and [23]). There have been many studies about record statistics. We refer to Westcott [67], Ahsanullah [2], [3], [4], [5], Nevzorov [56], Su et al. [61], and Ahsanullah and Nevzorov [6].

Since we deal with the ordered random variables it will be useful to mention some other models of ordered random variables, for example record values. The theory of record values is similar to the theory of order statistics.

Let $X_1, X_2, ...$ be a sequence of continuous iid random variables with cdf $F_X(x)$ and pdf $f_X(x)$. The rth upper record time $U(r)$ is defined as follows:

$$U(1) = 1,\ X_{U(1)} = X_1,$$

and

$$U(r+1) = \min\left\{i : X_i > X_{U(r)}\right\}, r = 1, 2, \dots\ .$$

Note that the rth record value is denoted by $X_{U(r)}$.

Then the pmf of the rth upper record time $U(r)$ is

$$P\left\{U(r) = k\right\} = \frac{|S(r-1, k-1)|}{r!},\ r = 1, 2, ..., k \text{ and } k = 1, 2, ..., \tag{2.20}$$

where $S(n,k)$ is Stirling numbers of second kind, i.e.,

$$\begin{aligned} S(n,k) &= \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \frac{i^n}{k!} \\ &= \sum_{i=0}^{k} (-1)^{k-i} \frac{i^n}{i!\,(k-i)!},\ 0 \le k \le n \text{ and } n = 0, 1, 2, \ldots . \end{aligned}$$

Indeed, $S(n,k)$ is the coefficient of the following expansion (see Balakrishnan and Nevzorov, [21])

$$x^n = \sum_{k=0}^{n} S(n,k) x(x-1)(x-2)\cdots(x-k+1)\,,\ n = 0, 1, 2, \ldots .$$

As $k \to \infty$, Westcott [67] showed that

$$P\{U(r) = k\} \sim \frac{(\log k)^{r-2}}{k^2 (r-1)!}.$$

For any $r = 1, 2, \ldots$ and any integers $1 = k_1 < k_2 < \cdots < k_r$, the joint pmf of the rth record time is given by

$$P\{U(1) = k_1, U(2) = k_2, \ldots, U(r) = k_r\} = \frac{1}{k_r} \prod_{i=2}^{r} \frac{1}{k_i - 1}.$$

The cdf of the rth upper record value is then

$$\begin{aligned} F_{X_{U(r)}}(x) &= P\left\{X_{U(r)} \le x\right\} \\ &= \frac{1}{(r-1)!} \int_0^{-\log(1-F(x))} z^{r-1} e^{-z} dz, \quad -\infty < x < \infty,\ r = 1, 2, \ldots \end{aligned} \tag{2.21}$$

For $r < s$, the joint cdf of the rth and sth record values is

$$f_{X_{U(r)},X_{U(s)}}(x_1,x_2)=\begin{cases}\frac{1}{\Gamma(r)\Gamma(s-r)}h(x_1)f(x_2) \\ \times\left[H(x_1)\right]^{r-1}\left[H(x_2)-H(x_1)\right]^{s-r-1}, & x_1\leq x_2 \\ 0, & \text{otherwise}\end{cases} \tag{2.22}$$

where $h(x)$ is the hazard function i.e.,

$$h(x)=\frac{dH_X(x)}{dx}=\frac{f_X(x)}{1-F_X(x)}$$

and

$$H(x)=-\ln(1-F_X(x)).$$

The joint pdf of the first n upper record values is

$$f_{X_{U(1)},X_{U(2)},\dots,X_{U(n)}}(x_1,x_2,...,x_n)=\begin{cases}f_X(x_n)\prod_{i=1}^{n-1}\frac{f_X(x_i)}{1-F_X(x_i)}, & x_1\leq x_2\leq\cdots\leq x_n \\ 0, & \text{otherwise}\end{cases} \tag{2.23}$$

Chapter 3

Asymptotic theory of exceedance statistics based on ordinary order statistics and records

The random threshold models based on ordered random variables have been a subject of many research published in statistical and engineering literature papers. Katzenbeiser [45], [46] was the first author who considered exceedance statistics in nonparametric hypothesis testing of equality of distributions. The limiting distributions of some placement statistics defined for order statistics were derived by Matveychuk and Petunin [52]. Exceedance models have also been considered for record values of iid random sequences. Bairamov [9] derived finite and limiting distributions of some exceedance statistics defined for record values of iid random sequences. Under some conditions coming out from the real life problems restrictions requiring independence and identically distributed random variables are avoidable. Under the condition that the underlying distribution of considered random sequence has points of discontinuity, Bairamov and Kotz [11] investigated the distributions of exceedance statistics based on record values and derived the distribution of the second record value. More general results concerning the distributions having countable number of points of discontinuity are presented in Bairamov and Khan [13]. Bairamov and Eryilmaz [15] studied the joint behavior of exceedances and precedences in a random threshold model. For more results on asymptotic distributions of exceedance statistics based on order statistics and record values and their applications we refer to Stepanov [60], Eryilmaz et al. [37], and Turhan [62]. Recently, Kemalbay and Bayramoglu [47] considered the bivariate random sample in exceedance model and derived the joint distribution of the ranks of order statistics.

In this chapter, applications of exceedance statistics of invariant confidence interval, exceedance statistics of inid random variables, exceedance statistics based on order statistics, concomitants and record statistics will be reviewed.

3.1 Invariant confidence intervals

Let $X_1, X_2, ..., X_n$ be the observed values of the random variable X with distribution function $F \in \Im$, where $\Im$ is a class of distribution functions and X_{n+1} be the future observation of random variable X. Let $h_1(x_1, x_2, ..., x_n)$ and $h_2(x_1, x_2, ..., x_n)$ be two real valued functions of the n variables such that $h_1(x_1, x_2, ..., x_n) \leq h_2(x_1, x_2, ..., x_n)$, $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$. Then for $\beta \in (0, 1)$, the random interval $W_\beta = (h_1(X_1, X_2, ..., X_n), h_2(X_1, X_2, ..., X_n))$ is called the invariant confidence interval for class $\Im$ containing the future observations with confidence level β if

$$P\left\{X_{n+1} \in W_\beta\right\} = \beta, \ \forall F \in \Im.$$

In the cases $h_1(X_1, X_2, ..., X_n) = -\infty$ or $h_2(X_1, X_2, ..., X_n) = \infty$, the invariant confidence interval W_β is said to be one sided. In other cases, that is $h_1(X_1, X_2, ..., X_n) \neq -\infty$ and $h_1(X_1, X_2, ..., X_n) \neq \infty$, W_β is called a two sided invariant confidence interval for X_{n+1}.

Let F be a continuous distribution function and $\Im$ a class of continuous distribution functions containing F. Assuming $h_1(X_1, X_2, ..., X_n) = X_{r:n}$ and $h_2(X_1, X_2, ..., X_n) = X_{s:n}$, where $1 \leq r < s \leq n$, Bairamov and Petunin [8] showed that for the class of all continuous functions $\Im = \Im_c$, the confidence level of the invariant confidence interval W_β is

$$P\{X_{n+1} \in W_\beta\} = \frac{s-r}{n+1}. \tag{3.1}$$

Furthermore, if $h_1(x_1, x_2, ..., x_n)$ and $h_2(x_1, x_2, ..., x_n)$ are continuous and symmetric functions, only the order statistics generate the invariant confidence interval W_β. Consider the observed values of continuous random variables $X_1, X_2, ..., X_n$ with cdf $F \in \Im$. Let $X_{n+1}, X_{n+2}, ..., X_{n+m}$ be the future observations of random variable X. Then, it is also true that

$$P\{X_{n+1}, X_{n+2}, ..., X_{n+m} \in W_\beta\} = \frac{n!(m+s-r-1)!}{(s-r-1)!(m+n)!}.$$

For more details on invariant confidence intervals, see Bairamov and Petunin [8] and Bairamov et al. [10].

3.2 Exceedance statistics of inid variables

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be a sequence of independent but not necessarily identically distributed random variables with continuous cdf $F_1(x), F_2(x), ...,$ $F_n(x)$, respectively and $\mathbf{Y} = (Y_1, Y_2, ..., Y_m)$ be another sequence of independent and identically distributed random variables with continuous cdf $G(y)$. Assume that $\mathbf{X} = (X_1, X_2, ..., X_n)$ and $\mathbf{Y} = (Y_1, Y_2, ..., Y_m)$ are independent. Let $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ denote the order statistics based on sample $\mathbf{X}$. Then for $i = 1, 2, ..., n$ and $1 \leq r < s \leq n$, define the following binary random variables

$$\eta_i = \begin{cases} 1, & \text{if } Y_i \in (X_{r:n}, X_{s:n}) \\ 0, & \text{otherwise} \end{cases}$$

It is obvious that the random variables η_i, $i = 1, 2, ..., n$, are dependent. Let $S_m = \sum_{i=1}^{m} \eta_i$. Then the statistics S_m counts the number of observations falling into random interval $(X_{r:n}, X_{s:n})$.

Essentially, derivation of the finite distribution of S_m encounters some problems due to the complexity of expressions of the joint pdf of rth and sth order statistics for inid random variables. It is given by

$$
\begin{aligned}
&P\{S_m = l\} = \sum_{i_1, i_2, \ldots, i_m} P\left\{\eta_{i_1} = 1, \ldots, \eta_{i_k} = 1, \eta_{i_{k+1}} = 0, \ldots, \eta_{i_m} = 0\right\} \\
&= \sum_{i_1, i_2, \ldots, i_m} P\{Y_{i_1} \in (X_{r:n}, X_{s:n}), \ldots, Y_{i_k} \in (X_{r:n}, X_{s:n}), \\
&Y_{i_{k+1}} \notin (X_{r:n}, X_{s:n}), \ldots, Y_{i_m} \notin (X_{r:n}, X_{s:n})\} \\
&= \sum_{i_1, i_2, \ldots, i_m} \iint P\{Y_{i_1} \in (x_1, x_2), \ldots, Y_{i_k} \in (x_1, x_2), \\
&Y_{i_{k+1}} \notin (x_1, x_2), \ldots, Y_{i_m} \notin (x_1, x_2)\} f_{X_{r:n}, X_{s:n}}(x_1, x_2) dx_1 dx_2 \qquad (3.2)
\end{aligned}
$$

When equation (2.8) is plugged into equation (3.2), the obtained equation is very complicated and is not useful for the application. Consequently, Bayramoglu and Giner [22] derived the asymptotic distribution of the statistics $\frac{S_m}{m}$ by the use of functionals for empirical distribution functions. They showed that

$$
\lim_{m \to \infty} \sup_{0 \le x \le 1} \left\{ \left| P\left\{ \frac{S_m}{m} \le x \right\} - P\{Q_{r,s} \le x\} \right| \right\} = 0,
$$

where $Q_{r,s} = G(X_{s:n}) - G(X_{r:n})$. Obtaining the general form of the finite distribution of $Q_{r,s}$ is not easy. Therefore, for some selected values of r and s the distributions of $Q_{r,s}$ are derived by Bayramoglu and Giner [22].

3.3 Exceedance statistics of iid random variables

In this section, we present an extensive literature review of exceedance statistics for independent and identically distributed random variables, based on order statistics, record statistics and concomitants.

3.3.1 Exceedance statistics based on order statistics

Let $X_1, X_2, ..., X_m$ be a sequence of iid random variables with common cdf $F_X(x)$ and $X_{m+1}, X_{m+2}, \dots, X_{m+n}$ be the future observations belonging to the same population. Assume that $X_1, X_2, ..., X_m$ and $X_{m+1}, X_{m+2}, ..., X_{m+n}$ are independent. For $1 \leq r \leq m$ denote the rth order statistic by $X_{r:m}$. Then define the following random variable

$$T_n(r) = \sum_{i=1}^{n} I_{(X_{r:m},\infty)}(X_{m+i}),$$

where $I_{[X_{r:m},\infty)}(X_{m+i}) = \begin{cases} 1, & if\ X_{m+i} > X_{r:m} \\ 0, & \text{otherwise} \end{cases} \quad i = 1, 2, ..., n.$

It is obvious that $T_n(r)$ is the total number of future observations greater than $X_{r:m}$. Then Gumbel and Von Schelling [41] derive the probability mass function of $T_n(r)$ given by

$$P\{T_n(r) = x\} = \frac{r\binom{m}{r}\binom{n}{x}}{(m+n)\binom{m+n-1}{n+x-1}},\ 1 \leq r \leq m \text{ and } 0 \leq x \leq n.$$

Let X be a random variable with cdf $F_X(x)$ and $Y_1, Y_2, ..., Y_m, ...$ is a sequence of iid random variables with a continuous cdf $F_Y(y)$. Assume that X is independent of $Y_1, Y_2, ..., Y_m, ...$. The survival function of $F_X(x)$ is defined by $\overline{F}_X(x) = P(X > x) = 1 - F_X(x)$. First, consider the sample $Y_1, Y_2, ..., Y_m$ and define the binary random variable I_i for $1 \le i \le m$ as

$$I_i = \begin{cases} 1, & \text{if } Y_i \le X \\ 0, & \text{otherwise} \end{cases}$$

If $S_m = \sum_{i=1}^{m} I_i$, then the statistic S_m counts the total number of observations which are less than or equal to X. For any integer $m \ge 1$, Wesolowski and Ahsanullah [66] derived the finite sample distribution of the statistic S_m, expected value, and variance as follows:

$$\begin{aligned} P\{S_m = i\} &= \binom{m}{i} E(F_Y^i(X)\overline{F}_Y^{m-i}(X)), \;\; i = 0, 1, ..., m \\ E(S_m) &= mE(F_Y(X)) \\ Var(S_m) &= mE(F_Y(X)\overline{F}_Y(X)) + m^2 Var(F_Y(X)). \end{aligned}$$

In order to derive the asymptotic distribution of S_m, they use characteristic functions and obtain

$$\lim_{m\to\infty} \sup_{0<x<1} \left| P\left\{ \frac{S_m}{m} \le x \right\} - P\{F_Y(X) \le x\} \right| = 0.$$

Wesolowski and Ahsanullah [66] also considered the number of the random variables $Y_1, Y_2, ..., Y_m, ...$ that are not exceeding the random threshold X. For any integer $m \ge 1$, define the random variable Z_m as

$$Z_m = \min\{i \ge 0 : \;\; Y_{i+1:m+i} > X\}.$$

The exact distribution, expected value, and variance Z_m are given as follows:

$$P\left\{Z_m = i\right\} = \binom{m+i-1}{m-1} E(F_Y^i(X)\overline{F}_Y^m(X)),\ i = 0, 1, ...,$$
$$E(Z_m) = mE\left(\frac{F_Y(X)}{\overline{F}_Y(X)}\right) I_{(0,1)}(F_Y(X)),$$
$$Var(Z_m) = mE\left(\frac{F_Y(X)}{F_Y^2(X)} I_{(0,1)}(F_Y(X))\right) + m^2 Var\left(\frac{F_Y(X)}{\overline{F}_Y(X)} I_{(0,1)}(F_Y(X))\right),$$

where

$$I_{(0,1)}(F_Y(X)) = \begin{cases} 1, & \text{if } \ 0 < F_Y(X) < 1 \\ 0, & \text{otherwise} \end{cases}$$

The asymptotic distribution of Z_m with the assumption $F_Y(X) < 1$ a.s is obtained as

$$\lim_{m\to\infty} \sup_{0<x<1} \left| P\left\{\frac{Z_m}{m} \leq x\right\} - P\left\{\frac{F_Y(X)}{\overline{F}_Y(X)} \leq x\right\}\right| = 0.$$

3.3.2 Exceedance statistics based on records

Let X be a random variable with cdf $F_X(x)$ and $Y_1, Y_2, ..., Y_m, ...$ a sequence of iid random variables with a continuous cdf $F_Y(y)$. Denote the survival function by $\overline{F}(.) = 1 - F(.)$. Assume that X is independent of $Y_1, Y_2, ..., Y_m, ...$.

Let $U(r)$ be the rth record of the random sample $Y_1, Y_2, ..., Y_m, ...$ and $U(1) = 1$. Then, Wesolowski and Ahsanullah [66] introduced the statistic K as follows

$$K = \min\left\{i \geq 0\colon\ Y_{U(i+1)} > X\right\}.$$

Obviously, the statistics K counts the total number of record values of $Y_1, Y_2, ..., Y_m, ...$ below X. Assume further that $P\left\{F_Y(x) < 1\right\} > 0$. Then the finite distributions of K is given by

$$P\left\{K = i\right\} = \frac{1}{i!}E(F_Y(X)(-\log(\overline{F}_Y(X)))^i I_{(0,1)}(F_Y(X))),\ i = 0, 1, ... \text{.}$$

Moreover, the expected value and variance of K have been calculated by

$$\begin{aligned} E(K) &= -E(-\log(\overline{F}_Y(X)))I_{(0,1)}F_Y(X) \\ Var(K) &= -E(-\log(\overline{F}_Y(X)))I_{(0,1)}F_Y(X)\ + Var(\log(\overline{F}_Y(X)))I_{(0,1)}(F_Y(X)). \end{aligned}$$

3.3.2.1 Distribution free properties of exceedance statistics based on records

Let $\mathbf{X}_\infty = (X_{1,}X_2, ..., X_n, ...)$ be a sequence of iid random variables with continuous cdf F and $\mathbf{Y}_\infty = (Y_1, Y_2, ..., Y_n, ...)$ be another sequence of iid random variables with the same cdf F. Assume that $\mathbf{X}_\infty$ and $\mathbf{Y}_\infty$ are independent. Consider the record values $X_{U(1)}, X_{U(2)}, ..., X_{U(n)}, ...$ based on the sample $\mathbf{X}_\infty$. Then, for $j = 1, 2, ..., m$ and $r = 1, 2, ...$ it is true that

$$P\left\{Y_j < X_{U(r)}\right\} = 1 - \frac{1}{2^r}.$$

Indeed,

$$
\begin{aligned}
P\left\{Y_j < X_{U(r)}\right\} &= \frac{1}{(r-1)!}\int_{-\infty}^{\infty} F(u)\left[\ln\frac{1}{1-F(u)}\right]^{r-1} dF(u)\\
&= \frac{1}{(r-1)!}\int_{0}^{1} u\left[\ln\frac{1}{1-u}\right]^{r-1} du\\
&= \frac{1}{(r-1)!}\int_{0}^{1} t^{r-1}(1-e^{-t})e^{-t}dt\\
&= 1-\frac{1}{2^r}.
\end{aligned}
$$

It is obvious that for any $j = 1, 2, ..., m$, $r = 1, 2, ...,$ and $r < s$

$$
P\left\{X_{U(r)} < Y_j < X_{U(s)}\right\} = \frac{1}{2^r} - \frac{1}{2^s}.
$$

Bairamov [9] defined the following binary random variable for $i = 1, 2, ..., m$ and $r = 1, 2, ...$

$$
\eta_i(r) = \begin{cases} 1, & \text{if } Y_i < X_{U(r)} \\ 0, & \text{otherwise} \end{cases}
$$

Denoting by $S_m(r) = \sum_{i=1}^{m} \eta_i(r)$, it is clear that $S_m(r)$ counts the total number of observations $Y_1, Y_2, ..., Y_n, ...$ which are less than $X_{U(r)}$.

The finite distribution of $S_m(r)$ is

$$
P\left\{S_m(r) = t\right\} = \frac{\binom{m}{t}}{(r-1)!}\int_{0}^{\infty} e^{-z(m-t+1)}(1-e^{-z})^t z^{r-1} dz.
$$

Furthermore, the expected value and variance of $S_m(r)$ are

$$E(S_m(r)) = m\left(1 - \frac{1}{2^r}\right),$$
$$Var(S_m(r)) = m^2\left(\frac{1}{3^r} - \frac{1}{2^{2r}}\right) + m\left(\frac{1}{2^r} - \frac{1}{3^r}\right).$$

The asymptotic distribution of $\frac{S_m(r)}{m}$ is

$$\lim_{m\to\infty}\sup_{0\le x\le 1}\left|P\left\{\frac{S_m(r)}{m}\le x\right\} - \frac{1}{(r-1)!}\int_0^x\left[\ln\left(\frac{1}{1-u}\right)\right]^{r-1}du\right| = 0.$$

Bairamov [9] also defined a new statistics $S_m^*(r) = \frac{S_m(r)-E(S_m(r))}{\sqrt{Var(S_m(r))}}$. Then $E(S_m^*(r)) = 0$, $Var(S_m^*(r)) = 1$, and the asymptotic distribution of $S_m^*(r)$ is

$$\lim_{m\to\infty}\sup_{\frac{a-1}{b}\le x\le\frac{a}{b}}\left|P\left\{\frac{S_m(r)}{m}\le x\right\} - \frac{b}{(r-1)!}\int_{\frac{a-1}{b}}^x\left[\ln\frac{1}{a-bu}\right]^{r-1}du\right| = 0,$$

where $a = \frac{1}{2^r}$ and $b = \sqrt{\frac{1}{3^r} - \frac{1}{2^{2r}}}$.

3.3.3 Exceedance statistics based on concomitants

Let $Z_1 = \{(X_i, Y_i),\ i = 1, 2, ..., n\}$ be a sequence of iid random variables with a continuous cdf $F(x, y)$ and $F_X(x)$, $F_Y(y)$ be the marginal cdf's of X and Y, respectively. Let $X_{r:n}$ denote the rth order statistics based on sample Z_1 and by $Y_{[r:n]}$ its concomitant, respectively. Consider another sequence of random variables $Z_2 = \{(X_{n+j}, Y_{n+j}),\ j = 1, 2, ..., m\}$ with continuous cdf $G(x, y)$. Let

$G_X(x)$ and $G_Y(y)$ be the corresponding marginal cdf's of X and Y, respectively. Define the following random variables

$$\nu_1 = \begin{cases} 1, & \text{if } X_{r:n} > X_{n+i} \\ 0, & \text{otherwise} \end{cases}$$

and

$$\nu_2 = \begin{cases} 1, & \text{if } X_{[r:n]} > X_{n+i} \\ 0, & \text{otherwise} \end{cases}$$

If $F_X(x)$ and $G_X(x)$ are continuous cdf's and $F_X(x) = G_X(x)$ for all $x \in \mathbb{R}$, then the finite sample distribution of ν_1 is given by

$$P\{\nu_1 = i\} = \frac{\binom{r+i-1}{r-1}\binom{m+n-i-r}{n-r}}{\binom{m+n}{n}}, \; i = 0, 1, ..., m.$$

Bairamov and Kotz [11] derived the finite distribution of ν_1 for arbitrary distributions having finite number of points of discontinuity.

Furthermore, the finite sample distribution of ν_2 is given by

$$P\{\nu_2 = j\} = \binom{m}{j-1}\frac{1}{Beta(r, n-r+1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big[(G_Y(y))^{j-1}(1-G_Y(y))^{m-j+1} \\ \times (F_X(x))^{r-1}(1-F_X(x))^{n-r} f(x,y)\Big]\,dxdy, \; j = 1, 2, ..., m+1,$$

where $Beta(r, n-r+1)$ is a Beta function.

3.3.4 Exceedance statistics for a sequence of dependent random variables

In this section, we present some results on exceedance statistics for the sequence of random variables whose finite dimensional distributions are n-variate Farlie-Gumbel-Morgenstein (FGM) distribution. The considered random variables having n-dimensional FGM distribution are dependent. It is also true that these random variables are asymptotically independent.

3.3.4.1 The case of Farlie-Gumbel-Morgenstein distribution

Definition 3.3.1. Let $\mathbf{X} = (X_1, X_2, ..., X_n, ...)$ be a random sequence with cdf's $F_i(x)$ and survival functions $\overline{F}_i(x) = 1 - F_i(x)$, for $i \geq 1$. Let $\beta(k,l)$ be a symmetric function, i.e., $\beta(k,l) = \beta(l,k), \forall k, l$. If for any permutations $i_1, i_2, ..., i_n$ of $1, 2, ..., n$ the joint cdf of $(X_{i_1}, X_{i_2}, ..., X_{i_n})$ is

$$F_{i_1,i_2,...,i_n}(x_{i_1}, x_{i_2}, ..., x_{i_n}) = \prod_{j=1}^{n} F_{i_j}(x_j) \left\{ 1 + \sum_{1 \leq l \leq k \leq n} \beta(k,l) \overline{F}_{i_k}(x_k) \overline{F}_{i_l}(x_l) \right\}, \tag{3.3}$$

then $\mathbf{X} = (X_1, X_2, ..., X_n, ...)$ is called the FGM sequence.

Definition 3.3.2. Consider the random sequence $\mathbf{X} = (X_1, X_2, ..., X_n, ...)$ with cdf $F_i(x)$ and survival functions $\overline{F}_i(x) = 1 - F_i(x)$, for $i \geq 1$. Assume that β_n is a sequence of real numbers. If the joint cdf of $X_1, X_2, ..., X_n$ is

$$F_{1,2,...,n}(x_1, x_2, ..., x_n) = \prod_{j=1}^{n} F_j(x_j) \left\{ 1 + \beta_n \sum_{1 \leq l \leq k \leq n} \overline{F}_k(x_k) \overline{F}_l(x_l) \right\},$$

then, $\mathbf{X} = (X_1, X_2, ..., X_n, ...)$ is called a simple-FGM (s-FGM) sequence.

Bairamov and Eryilmaz [12] showed that the admissible range of β_n, $n \geq 1$, for the n-variate distribution is

$$-\frac{1}{\binom{n}{2}} \leq \beta_n \leq \frac{1}{\left[\frac{n}{2}\right]},$$

where $[a]$ denotes the integer part of $a \in \mathbb{R}$. It can be easily seen that $X_1, X_2, ..., X_n, ...$ are asymptotically independent as $n \to \infty$.

Let Y be a random variable with cdf $G_Y(y)$ and $\mathbf{X} = (X_1, X_2, ..., X_m)$ be a s-FGM sequence with marginal cdf $F_i(x) = F(x)$ and pdfs $f_i(x) = f(x)$, $1 \leq i \leq m$ and $-\infty < x < \infty$. Consider a random variable Y independent from $\mathbf{X} = (X_1, X_2, ..., X_m)$. The joint pdf of $\mathbf{X} = (X_1, X_2, ..., X_m)$ is then

$$f_{1,2,...,m}(x_1, x_2, ..., x_m) = \prod_{j=1}^{m} f(x_j) \left\{ 1 + \beta_m \sum_{1 \leq l \leq k \leq m} (1 - 2F(x_k))(1 - 2F(x_l)) \right\},$$

$$-\infty < x_1, x_2 ..., x_m < \infty \text{ and } -\frac{1}{\binom{m}{2}} \leq \beta_m \leq \frac{1}{\left[\frac{m}{2}\right]}.$$

For $1 \leq j \leq m$ define the binary random variable ξ_j as follows:

$$\xi_j = \begin{cases} 1, & \text{if } X_j \leq Y \\ 0, & \text{otherwise} \end{cases}$$

Consider $S_m = \sum_{j=1}^{m} \xi_j$. Here, S_m counts the number of observations of X_j' s not exceeding the random threshold Y. Then for $k = 0, 1, ..., m$, the finite sample distribution of S_m is given by

$$
\begin{aligned}
P\left\{S_m=k\right\} = \binom{m}{k} \Big[& E(F^k(Y)\overline{F}^{m-k}(Y)) \\
& + \beta_m \left(\frac{k(k-1)}{2} E(F^k(Y)\overline{F}^{m-k+2}(Y)) - k(m-k)E(F^{k+1}(Y)\overline{F}^{m-k+1}(Y)) \right. \\
& \left. + \frac{(m-k)\,(m-k-1)}{2} E(F^{k+2}(Y)\overline{F}^{m-k}(Y)) \right) \Big] \ ,
\end{aligned}
$$

where $\overline{F}(x)=1-F(x)$.

If $F=G$, then for $k=0,1,...,m$ and $-\frac{1}{\binom{m}{2}} \leq \beta_m \leq \frac{1}{\left[\frac{m}{2}\right]}$, the finite sample distribution of S_m is

$$
\begin{aligned}
P\left\{S_m=k\right\} = \binom{m}{k} \Big[& Beta(k+1,m-k+1) + \beta_m \left(\frac{k(k-1)}{2} Beta(k+1,m-k+3) \right. \\
& - k(m-k)Beta(k+2,m-k+2) \\
& \left. + \frac{(m-k)(m-k-1)}{2} Beta(k+3,m-k+1) \right) \Big] ,
\end{aligned}
$$

where $Beta(x,y)$ is a beta function.

Let $\mathbf{X}=(X_1,X_2,...,X_n)$ be a s-FGM sequence with marginal cdf $F(x)$ and $\mathbf{Y}=(Y_1,Y_2,...,Y_m,...)$ be another sequence of iid random variables with cdf $G(y)$. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics of $\mathbf{X}$. Bairamov and Eryilmaz [12] also introduced a binary random variable η_j, $1 \leq j \leq m$ and $1 \leq r \leq n$, based on the rth order statistics of the sequence $\mathbf{X}$ as

$$
\eta_j = \begin{cases} 1, & \text{if } Y_j \leq X_{r:n} \\ 0, & \text{otherwise} \end{cases}
$$

Define the statistic $S_m(r)$ as follows

$$S_m(r) = \sum_{j=1}^{m} \eta_j.$$

In this case $S_m(r)$ counts the number of observations Y_i' s falling into the interval $(-\infty, X_{r:n}]$.

If $F = G$, then for $k = 0, 1, ..., m$ and $-\frac{1}{\binom{n}{2}} \leq \beta_n \leq \frac{1}{\left[\frac{n}{2}\right]}$, the finite sample distribution of $S_m(r)$ is given by

$$
\begin{aligned}
P\left\{S_m(r) = k\right\} = \binom{m}{k} \sum_{t=r}^{n} (-1)^{s-r} \binom{s-1}{r-1} \binom{n}{s} & \left[tBeta(t+k, m-k+1) \right. \\
& + \beta_n \left(\frac{s^2(s-1)}{2} Beta(s+k, m-k+3) \right. \\
& \left. \left. - s(s-1) Beta(s+k+1, m-k+2) \right) \right].
\end{aligned}
$$

Chapter 4

Main Results

In this chapter, we aim to derive the exact and asymptotic distributions of the exceedance statistics defined for the bivariate random sequences $Z_1 = \{(X_i, Y_i),$ $i = 1, 2, ..., n\}$ and $Z_2 = \{(X_{n+j}, Y_{n+j}),\ j = 1, 2, ..., m, ...\}$. Here and thereafter, Z_1 and Z_2 are assumed to be independent. We also assume that X and Y are dependent random variables. Therefore, the distribution functions are expected to involve the copulas of distributions of the first and second sequences. Below, we provide an insight into the definition of copula function, since the theoretical results and examples presented in this thesis rely on the copulas.

Definition 4.0.1. (Nelsen, 2006, p. 10) A two dimensional Copula is a function $C : [0, 1]^2 \to [0, 1]$, such that

i For every $t, s \in [0, 1]$

$$C(t, 0) = C(0, s) = 0$$

and

$$C(t, 1) = t \text{ and } C(1, s) = s.$$

ii For every $t_1 \leq t_2$ and $s_1 \leq s_2$ in $[0, 1]$,

$$C(t_2, s_2) - C(t_1, s_2) - C(t_2, s_1) + C(t_1, s_1) \geq 0.$$

Theorem 4.1 (Sklar's Theorem) *(Nelsen, 2006, p. 17) Let X and Y be a pair of random variables with joint distribution function $H(x,y)$ and marginal distribution functions $F_X(x)$ and $F_Y(y)$, respectively. Then for all $x, y \in \mathbb{R} \cup \{-\infty, \infty\}$, there exists a copula function C such that*

$$H(x,y) = C(F_X(x), F_Y(y)). \tag{4.1}$$

If $F_X(x)$ and $F_Y(y)$ are continuous distribution functions, then the Copula function is unique. In the contrary case, if C is a copula and $F_X(x)$ and $F_Y(y)$ are distribution functions, then the function $H(x,y)$ is a joint distribution function with marginals $F_X(x)$ and $F_Y(y)$.

Definition 4.0.2 (Quasi-Inverse Function)**.** Let F be cdf of a random variable X. Then a quasi-inverse function of F is any function $F^{-1}(u) : [0,1] \to \mathbb{R} \cup \{-\infty, \infty\}$ such that:

$$F^{-1}(u) = \inf \{x : F(x) \geq u\} = \sup \{x : F(x) \leq u\}.$$

If F is a strictly increasing distribution function, then the quasi-inverse function F^{-1} is the ordinary inverse function.

Corollary 4.2 *Let $F_X(x)$ and $F_Y(y)$ be continuous continuous distribution functions of X and Y, respectively. Let $F_X^{-1}(x)$ and $F_Y^{-1}(y)$ be the quasi inverse functions of X and Y, respectively and $H(x,y)$ be the joint cdf of X and Y. Then*

$$C(u,v) = H(F_X^{-1}(x), F_Y^{-1}(y)). \tag{4.2}$$

Below, the definition of Archimedean family of copulas is given.

Definition 4.0.3 (Archimedean Copula)**.** (Nelsen, 2006, p. 110) Consider a continuous, strictly decreasing function $\phi : [0,1] \to [0,\infty]$ such that $\phi(1) = 0$ and $\phi(0) = \infty$. The class of copulas $C(u,v)$ of the form

$$C(u,v) = \phi^{-1}\left\{\phi(u) + \phi(v)\right\} \tag{4.3}$$

is called Archimedean family of copulas, and the function ϕ is called generator function.

Archimedean family of copulas have the following properties:

1. $C(u,v)$ is symmetric for $\forall u, v$ in $[0,1]$, i.e.,

$$C(u,v) = C(v,u), \forall (u,v) \in [0,1]^2$$

2. $C(u,v)$ is associative for $\forall u, v, w$ in $[0,1]$, i.e.,

$$C(C(u,v),w) = C(u, C(v,w));$$

3. If a is any constant, then $a\phi$ is a generator function too.

4.1 Exceedance statistics based on simple model

In this section, we present the exact and asymptotic distributions of exceedance statistics based on random threshold (X,Y). Furthermore, asymptotic distributions of normalized exceedance statistics are investigated. Examples and numerical results are provided for some well known copulas.

4.1.1 Finite distributions of exceedance statistics based on simple model

Assume that $Z_1 = (X, Y)$ is a bivariate random vector with absolutely continuous distribution function $F(x, y) = C_1(F_X(x), F_Y(y))$, where $C_1(u, v), (u, v) \in [0, 1]^2$, is a connected copula and $F_X(x)$, $F_Y(y)$ are the corresponding marginal distributions of X and Y, respectively and $f_X(x)$ and $f_Y(y)$ are the probability density functions of X and Y defined by $F'_X(x) = f_X(x)$ and $F'_Y(y) = f_Y(y)$. Furthermore, let $Z_2 = \{(X_i, Y_i),\ i = 1, 2, ..., m, ...\}$ be a sequence of independent random vectors with absolutely continuous distribution function
$G(x, y) = C_2(F_X(x), F_Y(y))$ and the same marginal distributions $F_X(x)$ and $F_Y(y)$, where C_2 is a connected copula. Assume that Z_1 and Z_2 are independent.

For $i = 1, 2, ..., m$ define the binary random variables

$$\xi_i \equiv I_M(X_i, Y_i) = \begin{cases} 1, & (X_i, Y_i) \in M \\ 0, & \text{otherwise} \end{cases}$$

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

Now define the random variable $S_m = \sum_{i=1}^{m} \xi_i$. It is clear that the exceedance statistics S_m counts the number of bivariate observations (X_i, Y_i), $i = 1, 2, ...,$ falling into the random set M and the random variables ξ_i, $i = 1, 2, ...,$ are dependent. In the following theorem, we obtain the finite sample distribution of S_m.

Theorem 4.3 *(See Erem and Bayramoglu [35]) It is true that*

$$P\{S_m = k\} = \binom{m}{k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G(x, y)]^k [1 - G(x, y)]^{m-k} dF_{X,Y}(x, y). \quad (4.4)$$

The expression in terms of copulas is

$$P\{S_m = k\} = \binom{m}{k} \int_0^1 \int_0^1 [C_2(u,v)]^k [1 - C_2(u,v)]^{m-k} dC_1(u,v), \tag{4.5}$$

where $C_2(u,v)$ is a copula corresponding to joint distribution function $G(x,y)$ and $C_1(u,v)$ is a copula corresponding to joint distribution function $F(x,y)$.

Proof. First we show that

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

Indeed, define

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

and observe that the complement of A_{i_j} as

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

For simplicity, we define the random set E as

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

Then conditioning with respect to $X = x$ and $Y = y$ we obtain

$$\begin{aligned} P\{S_m = k\} &= \sum_{i_1, i_2, \ldots, i_m} P\left\{A_{i_1} A_{i_2} \cdots A_{i_k} A_{i_{k+1}}^c \cdots A_{i_m}^c\right\} \\ &= \sum_{i_1, i_2, \ldots, i_m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P\{E \mid X = x, Y = y\} dF_{X,Y}(x,y) \\ &= \sum_{i_1, i_2, \ldots, i_m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P\left\{a_{i_1} a_{i_2} \cdots a_{i_k} a_{i_{k+1}}^c \cdots a_{i_m}^c\right\} dF_{X,Y}(x,y), \end{aligned} \tag{4.6}$$

where the sum $\sum_{i_1, i_2, \ldots, i_m}$ extends over all permutations of $i_1, i_2, ..., i_m \in \{1, 2, ..., m\}$,

$$a_{i_j} = \left\{X_{i_j} < x, Y_{i_j} < y\right\}, \text{ and } a_{i_j}^c \text{ is the complement of } a_{i_j}.$$

Since the events a_{i_j} are independent, the probability in (4.6) can be written as

$$P\left\{a_{i_1}a_{i_2}\cdots a_{i_k}a^c_{i_{k+1}}\cdots a^c_{i_m}\right\} = P(a_{i_1})P(a_{i_2})\cdots P(a_{i_k})P(a^c_{i_{k+1}})\cdots P(a^c_{i_m}).$$

Therefore, the finite sample distribution of S_m can be simplified as

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

Using the probability integral transformation $F_X(x) = u,\ F_Y(y) = v$ in equation (4.4), we obtain equation (4.5). Thus, the theorem is proved. □

Proposition 4.4 *It is clear that*

$$E(S_m) = mE(G(X,Y)) = mE(C_2(U,V)), \tag{4.7}$$

$$\begin{aligned} Var(S_m) &= mE(G(X,Y)) - mE(G^2(X,Y)) + m^2Var(G(X,Y)) \\ &= mE(C_2(U,V)) - mE(C_2^2(U,V)) + m^2Var(C_2(U,V)), \end{aligned} \tag{4.8}$$

where (U,V) is a random vector with Uniform(0,1) marginals having joint cdf $C_2(u,v)$.

Proof.

$$\begin{aligned} E(S_m) &= \sum_{k=0}^{m} kP\left\{S_m = k\right\} \\ &= \sum_{k=0}^{m} k\binom{m}{k}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[G(x,y)\right]^k\left[1-G(x,y)\right]^{m-k} dF_{X,Y}\left(x,y\right). \\ &= \sum_{k=0}^{m} k\binom{m}{k}E\left(G^k(X,Y)(1-G(X,Y))^{m-k}\right) \\ &= E\left[\sum_{k=0}^{m} k\binom{m}{k}G^k(X,Y)(1-G(X,Y))^{m-k}\right]. \end{aligned} \tag{4.9}$$

Consider the binomial random variable Z with parameters m and T. It is clear that $E(Z) = mT$ and $Var(Z) = mT(1-T)$. Denote $T \equiv G(X,Y)$. Then one can write equation (4.9) in the following form

$$
\begin{aligned}
&= E\left[\sum_{k=0}^{m} k\binom{m}{k}T^{k}(1-T)^{m-k}\right] \\
&= E(E(Z)) \\
&= E(mT) \\
&= mE(G(X,Y)) \qquad (4.10)
\end{aligned}
$$

Similarly, the second moment of S_m is

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

Hence,

$$
\begin{aligned}
Var(S_m) &= E(S_m^2) - (E(S_m))^2 \\
&= m^2E(G^2(X,Y)) + mE\left[G(X,Y)(1-G(X,Y))\right] \\
&\quad - m^2\left[E(G(X,Y))\right]^2 \\
&= m^2(E(G^2(X,Y)) - \left[E(G(X,Y))\right]^2) \\
&\quad + mE(G(X,Y)) - mE(G^2(X,Y)) \\
&= mE(G(X,Y)) - mE(G^2(X,Y)) + m^2Var(G(X,Y)). \qquad (4.11)
\end{aligned}
$$

Using probability integral transformation in equations (4.10) and (4.11), the proof is completed. □

Example 4.1 (Product Copula)**.** Consider the trivial case where the random variables X and Y are independent and so are X_i and Y_i. Then
$C_1(u,v) = C_2(u,v) = uv$.

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

and finally,

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

Example 4.2 (Gumbel-Barnett Copula)**.** Let $C_2(u,v) = uv\exp(-\theta \ln u \ln v)$, $\theta \in (0,1]$ be a Gumbel-Barnett family of copulas and $C_1(u,v) = uv$. The Gumbel-Barnet family of copulas belongs to the class of Archimedean copulas with generating function $\varphi_\theta(t) = \ln(1-\theta\ln(t))$ (see Nelsen, 2006, p. 119). It is obvious that $\theta = 0$ implies independency. For this case the finite sample distribution of S_m is

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

As it is seen from example 4.2, the pmf of S_m involves sum and integral of exponential and logarithmic functions. Therefore, it can be easily calculated numerically for different values of m and k.

4.1.1.1 Numerical results

In Table 4.1, some numerical values for probability $P\{S_m = k\}$ for $m = 5$ are provided for the following cases:

Case i: $C_1(u,v) = C_2(u,v) = uv$.

Case ii: $C_1(u,v) = uv$, $C_2(u,v) = uv\exp(-\theta \ln u \ln v)$, $\theta \in (0,1]$.

Case iii: $C_1(u,v) = uv$, $C_2(u,v) = [\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-1/\theta}$, $\theta \in [-1,\infty)\backslash\{0\}$, in this case $C_2(u,v)$ is called Clayton copula which corresponds to the bivariate Pareto distribution. Clayton copula is also a well-known survival copula of Archimedean family copulas with generating function $\varphi_\theta(x) = \frac{1}{\theta}\left(x^{-\theta} - 1\right)$. $\alpha \to 0$ implies independency.

Case iv: $C_1(u,v) = uv$, $C_2(u,v) = \frac{uv}{1-\theta(1-u)}$, $\theta \in [-1,0)$. $C_2(u,v)$ is known as Ali–Mikhail–Haq copula and it is a member of Archimedean family copulas with generating function $\varphi_\theta(x) = \ln\frac{1-\theta(1-x)}{x}$ (Nelsen, 2006; p. 116 and Balakrishnan and Lai, 2009; p. 90). It is obvious that for $\theta = 0$, $C_2(u,v)$ is independent.

	Case i	Case ii θ=0.5	Case iii θ=1	Case iv θ=-1
	$P\{S_5 = k\}$	$P\{S_5 = k\}$	$P\{S_5 = k\}$	$P\{S_5 = k\}$
k = 0	0.408	0.269	0.327	0.458
k = 1	0.242	0.179	0.259	0.337
k = 2	0.158	0.154	0.187	0.154
k = 3	0.103	0.141	0.124	0.0436
k = 4	0.061	0.132	0.072	0.0069
k = 5	0.028	0.125	0.031	0.0005

Table 4.1: Exact distribution of S_m for different copulas

As it is seen from Table 4.1, the probability mass function of S_m decreases as k increases.

4.1.2 Asymptotic distributions of exceedance statistics in simple model

In this section, we are going to obtain the asymptotic distribution of $\frac{S_m}{m}$ when $m \to \infty$. The following theorem presents a result for $\lim\limits_{m\to\infty} P\left\{\frac{S_m}{m} \leq x\right\}$.

Theorem 4.5 *(See Erem and Bayramoglu [35]) It is true that the statistics* $\frac{S_m}{m}$ *has a continuous limiting distribution*

$$W(x) = \begin{cases} 0, & x < 0 \\ P\left\{C_2(U,V) \leq x\right\}, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases} \tag{4.12}$$

where (U, V) is a bivariate random vector with copula $C_2(u, v)$. Furthermore, the cdf $W(x)$ can be also written in the following form

$$W(x) = \begin{cases} 0, & x < 0 \\ P\left\{G(X,Y) \leq x\right\}, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}. \tag{4.13}$$

To prove the theorem, we need the following lemma.

Lemma 4.6 *Let* $G_m^*(u,v)$ *be a bivariate empirical distribution function and* $g_{X,Y}(u,v)$ *is a continuous function defined on* $[-A, A] \times [-B, B]$. *Then for any partition* $-A = x_0 < x_1 < x_2 < \cdots < x_n = A$ *and* $-B = y_0 < y_1 < y_2 < \cdots < y_n = B$, *it is true that*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{X,Y}(u,v)\, dG_m^*(u,v) = \frac{1}{m}\sum_{i=1}^{m} g(X_i, Y_i).$$

Proof of Lemma 4.6. Observe that from the definition of the Stieltjes integral of a real function $g(u, v)$ with respect to empirical distribution function $G_m^*(u,v)$, we have

$$\lim_{A,B\to\infty}\int_{-A}^{A}\int_{-B}^{B} g(u,v)G_m^*(u,v) = \frac{1}{m}\sum_{i=1}^{m} g(X_i,Y_i).$$

Indeed, for any partition $-A = x_0 < x_1 < x_2 < \cdots < x_n = A$ and $-B = y_0 < y_1 < y_2 < \cdots < y_n = B$, the Stieltjes integral of $g(u,v)$ with respect to $G_m^*(u,v)$ is

$$\sum_{i=1}^{n}\sum_{j=1}^{n} g(\xi_i,\eta_j)[G_m^*(x_i,y_j) - G_m^*(x_i,y_{j-1}) - G_m^*(x_{i-1},y_j) + G_m^*(x_{i-1},y_{j-1})],$$

where $\xi_i \in (x_{i-1},x_i], \eta_j \in (y_{j-1},y_j]$, $i,j = 1,2,...,n$. Since

$$G_m^*(u,v) = \begin{cases} 0, & u < X_{1:m} \text{ or } v < Y_{1:m} \\ & \{X_{i:m} \le u < X_{i+1:m} \text{ and } Y_{i:m} \le v < Y_{i+1:m}\} \text{ or} \\ \frac{i}{m}, & \{u > X_{i+1:m} \text{ and } Y_{i:m} \le v < Y_{i+1:m}\} \text{ or} \\ & \{X_{i:m} \le u < X_{i+1:m} \text{ and } v > Y_{i+1:m}\},\ 1 \le i \le m \\ \cdots & \cdots \\ 1, & u > X_{m:m} \text{ and } v > Y_{m:m}, \end{cases}$$

it is clear that $G_m^*(X_{i:m},Y_{i:m}) - G_m^*(X_{i:m},Y_{i:m}-0) - G_m^*(X_{i:m}-0,Y_{i:m}) + G_m^*(X_{i:m}-0,Y_{i:m}-0) = \frac{i}{m} - \frac{i-1}{m} - \frac{i-1}{m} + \frac{i-1}{m} = \frac{1}{m}$ and $G_m^*(x_i,y_j) - G_m^*(x_i,y_j-0) - G_m^*(x_i-0,y_j) + G_m^*(x_i-0,y_j-0) = 0$ for any other points x_i and y_j, where $G_m^*(X_{i:m},Y_{i:m}-0) = \lim\limits_{Y_{i:m}\to Y_{i:m}^-} G_m^*(X_{i:m},Y_{i:m})$. That is why

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{n} g(\xi_i,\eta_j)[G_m^*(x_i,y_j) - G_m^*(x_i,y_{j-1}) \\ &- G_m^*(x_{i-1},y_j) + G_m^*(x_{i-1},y_{j-1})] \\ &= \frac{1}{m}\sum_{i=1}^{m} g(X_{i:m},Y_{i:m}) = \frac{1}{m}\sum_{i=1}^{m} g(X_i,Y_i).\end{aligned}$$

The lemma is proved. □

Proof of Theorem 4.6. The probability of $\frac{S_m}{m}$ being less than or equal to a real number x can be written as

$$P\left\{\frac{S_m}{m} \le x\right\} = P\left\{\frac{1}{m}\sum_{i=1}^{m} I_{(-\infty,X]\times(-\infty,Y]}\left(X_i,Y_i\right) \le x\right\}$$
$$= P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{X,Y}\left(u,v\right) dG_m^*\left(u,v\right) \le x\right\}, \qquad (4.14)$$

where

$$I_A(x,y) = \begin{cases} 1, & (x,y) \in A \\ 0, & otherwise \end{cases},$$

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

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 on the basis of $(X_1,Y_1),...,(X_m,Y_m)$. Now, conditioning (4.14) on $X=x$ and $Y=y$ and using the independence of the random vectors (X_i,Y_i), $i=1,2,...,m,...$ and (X,Y) we obtain

$$P\left\{\frac{S_m}{m} \le x\right\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{X,Y}\left(u,v\right) dG_m^*\left(u,v\right) \le x \mid X=t, Y=s\right\} dF_{X,Y}(t,s)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{t,s}\left(u,v\right) dG_m^*\left(u,v\right) \le x\right\} dF_{X,Y}(t,s). \qquad (4.15)$$

Using the Glivenko-Cantelli Theorem (Borovkov, 1998; p. 5), we have

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

and the continuity property of the integral functional

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

yields

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

Then we can write

$$\begin{aligned}
W(x) &\equiv \lim_{m\to\infty} P\left\{\frac{S_m}{m} \le x\right\} \\
&= \lim_{m\to\infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{t,s}\left(u,v\right) dG_m^*\left(u,v\right) \le x\right\} dF_{X,Y}(t,s) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lim_{m\to\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{t,s}\left(u,v\right) dG_m^*\left(u,v\right) \le x\right\} dF_{X,Y}(t,s) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{t,s}\left(u,v\right) dG\left(u,v\right) \le x\right\} dF_{X,Y}(t,s) \\
&= P\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{X,Y}\left(u,v\right) dG\left(u,v\right) \le x\right\} \\
&= P\left\{\int_{-\infty}^{X}\int_{-\infty}^{Y} dG\left(u,v\right) \le x\right\} \\
&= P\{G(X,Y) \le x\}. \qquad (4.16)
\end{aligned}$$

Using the probability integral transformation $U = F_X(X)$ and $V = F_Y(Y)$ in (4.16), we obtain (4.12). Thus, the theorem is proved. □

Corollary 4.7 *If $C_2(u,v) = C_1(u,v) = C(u,v)$, then*

$$\lim_{m\to\infty} \sup_{-\infty<x<\infty} \left| P\left\{\frac{S_m}{m} \le x\right\} - P\{C(U,V) \le x\}\right| = 0, \qquad (4.17)$$

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

Note that in Corollary 4.7, in the case where $C_2(u,v) = C_1(u,v) = C(u,v)$, the function $P\{F(X,Y) \le x\}$ and $P\{C(U,V) \le x\}$ are the well known Kendall distribution function (see Nelsen et al. (2003) [54] and Cherubini et al. [29]). In Genest and Rivest [38] the Kendall distribution function for different Archimedean copulas are calculated. The following proposition by Genest and Rivest [38] paves the way for a method to calculate the Kendall distribution function for the class of Archimedean copulas.

Proposition 4.8 *(see Genest and Rivest [38]) Let X and Y be uniformly distributed random variables whose copula $C(x,y)$ is of the form $\phi^{-1}\{\phi(x)+\phi(y)\}$ for some convex decreasing function ϕ defined on $(0,1]$ satisfying $\phi(1)=0$. Set $U = \frac{\phi(X)}{\{\phi(X)+\phi(Y)\}}$, $V = C(X,Y)$, and $\lambda(v) = \frac{\phi(v)}{\phi'(v)}$ for $0 < v \le 1$. Then,*

(a) U is uniformly distributed on $(0,1)$,

(b) V is distributed as $K(v) = v - \lambda(v)$ on $(0,1)$, and

(c) U and V are independent random variables.

Below in the sequence of examples 4.3– 4.7, the Kendall distribution function of some well-known Archimedean family copulas are provided using Proposition 4.8. Actually, it follows from Corollary 4.7 that these Kendall distribution functions are the limiting distributions of $\frac{S_m}{m}$.

Example 4.3 (Product Copula)**.** Let $C(u,v)$ be the product copula, then the Kendall distribution function is

$$W(x) = x - x\log(x),\ 0 < x < 1.$$

Example 4.4 (Clayton Copula)**.** Let $C(u,v)$ be Clayton copula. Then,

$$W(x) = x + \frac{x(1-x^{\alpha})}{\alpha},\ \ 0 < x < 1.$$

For $\alpha = 1$,

$$W(x) = x + x(1-x),\ \ 0 < x < 1.$$

Example 4.5 (Frank Copula)**.** Let $C(u,v)$ be a Frank family copula, i.e.,

$$C(u,v) = -\frac{1}{\alpha}\ln\left(1+\frac{\left(e^{-\alpha u}-1\right)\left(e^{-\alpha v}-1\right)}{e^{-\alpha}-1}\right), \alpha\in\mathbb{R}\backslash\{0\}$$

with generating function $\varphi_\alpha(x) = \log\left(\frac{1-\exp(-\alpha)}{1-\exp(-\alpha x)}\right)$. The Frank family copulas are the only Archimedean copulas satisfying $C(u,v) = \widehat{C}(u,v)$, where $\widehat{C}(u,v)$ is the survival copula. As $\alpha\to\infty$, $C(u,v)$ becomes the product copula. Then

$$W(x) = x+\frac{1-\exp\left(-\alpha x\right)}{\alpha\exp\left(-\alpha x\right)}\log\left(\frac{1-\exp\left(-\alpha\right)}{1-\exp\left(-\alpha x\right)}\right),\ 0<x<1.$$

For $\alpha = 1$, the function $W(x)$ can be written as

$$W(x) = x+\frac{1-e^{-x}}{e^{-x}}\log\left(\frac{1-e^{-1}}{1-e^{-x}}\right),\ 0<x<1.$$

Example 4.6 (Gumbel Copula)**.** Let $C(u,v)$ be a Gumbel copula, i.e., $C(u,v) = e^{-[(-\ln u)^\alpha+(-\ln v)^\alpha]^{1/\alpha}}$, $\alpha\in[1,\infty)$ with generator function $\varphi_\alpha(x) = \{-\log(x)\}^\alpha$. Then the Kendall distribution function is given by

$$W(x) = x-\frac{x\log(x)}{\alpha+1},\ 0<x<1.$$

For $\alpha = 1$,

$$W(x) = x-\frac{x\log(x)}{2},\ 0<x<1.$$

It is also clear that $\alpha = 0$ indicates independency.

Example 4.7 (Log Copula)**.** Let $C(u,v)$ be the log copula, i.e.,

$$C(u,v) = \exp\left\{\alpha\gamma\left[1-\left\{\left(1-\frac{\ln u}{\alpha\gamma}\right)^{\alpha+1}+\left(1-\frac{\ln v}{\alpha\gamma}\right)^{\alpha+1}-1\right\}^{\frac{1}{\alpha+1}}\right]\right\},$$

$\alpha,\gamma\in(0,\infty)$,

with generating function $\varphi_{\alpha,\gamma}(x) = \{1-\log(x)/\alpha\gamma\}^{\alpha+1}-1$. Then

$$W(x) = x+\frac{\alpha\gamma x\left[\{1-\log(x)/\alpha\gamma\}^{\alpha+1}-1\right]}{(\alpha+1)\{1-\log(x)/\alpha\gamma\}^{\alpha}},\ 0<x<1.$$

For $\alpha=\gamma=1$, the function $W(x)$ can be written by

$$W(x) = x+\frac{x\left[(1-\log(x))^2-1\right]}{2(1-\log(x))},\ 0<x<1.$$

In Figure 4.1, the asymptotic distributions of $\frac{S_m}{m}$ for copulas given in examples 4.3, 4.4, 4.5, 4.6 and 4.7 are provided for $\alpha=1$. Note that $W_i(x)$, $i=1,2,...,5$ is the asymptotic distribution, for the independent, Clayton, Frank, Gumbel and Log Copula, respectively.

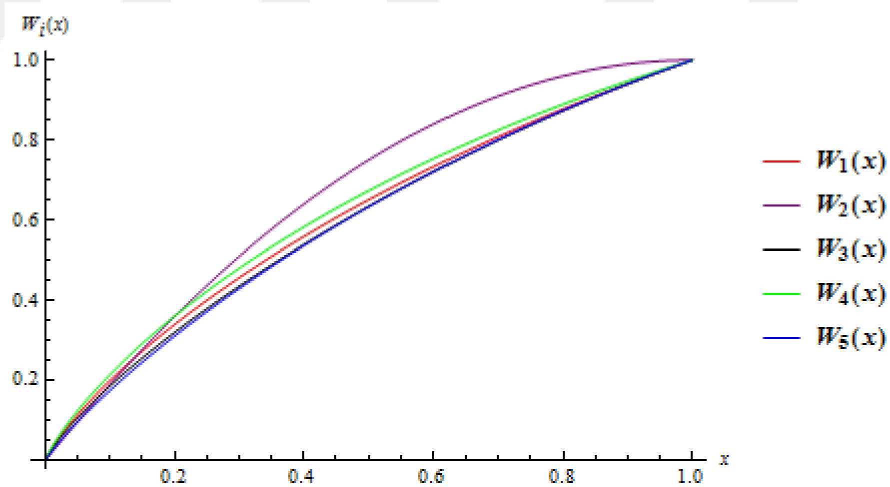

Figure 4.1: The graphs of the limiting distributions of $\frac{S_m}{m}$ for copulas given in examples 4.3, 4.4, 4.5, 4.6 and 4.7

Example 4.8 (FGM Copula)**.** Let $C_1(u,v)$ be a Farlie-Gumbel-Morgenstein (FGM) copula. FGM copulas are one of the well-known copulas and they have various applications in modelling survival data, reliability engineering, risk analysis and insurance (see Louzada et al. [51] and Danaher et al. [30]). The probability density function of the FGM copula is

$$c_1(u,v) = 1 + \alpha\,(1-2u)\,(1-2v)\,, -1 \leq \alpha \leq 1,$$

and the distribution function of the FGM copula is

$$C_1(u,v) = uv\left[1 + \alpha\,(1-u)\,(1-v)\right], -1 \leq \alpha \leq 1.$$

Lai [49] showed that U and V are positive quadrant dependent (PQD) for $0 \leq \alpha \leq 1$, i.e.,

$$\overline{F}(x,y) \geq \overline{F}_X(x)\overline{F}_Y(y),\ \forall (x,y) \tag{4.18}$$

or

$$C(u,v) \geq uv,\ \forall u,v \in [0,1]\ , \tag{4.19}$$

where $\overline{F}(x,y)$ is the survival function of (X,Y) and $\overline{F}_X(x)$ and $\overline{F}_Y(y)$ are the marginal survival functions of X and Y, respectively. For more details see Lai [49] and Lehmann [50].

Example 4.9. Let $C_1(u,v) = uv\left[1 + \alpha\,(1-u)\,(1-v)\right],\ -1 \leq \alpha \leq 1$ and $C_2(u,v) = uv$. Then

$$F_\alpha(x) \equiv P\left\{C_2(u,v) \leq x\right\} = \iint\limits_{\{(u,v):uv\leq x\}} c_1(u,v)dudv$$

After the transformation $t = uv$ and $s = u$ with Jacobian $|J| = \frac{1}{s}$, we have

$$F_\alpha(x) \equiv \int_0^x \int_t^1 \left[1 + \alpha\,(1-2s)\,(1-\frac{2t}{s})\right] \frac{1}{s} dsdt$$

$$F_\alpha(x) = \begin{cases} 0, & x < 0 \\ x\,(1 + 3\alpha\,(x-1) - (1 + \alpha + 2\alpha x)\ln x) & 0 \leq x \leq 1 \\ 1 & x > 1, \end{cases}$$

and

$$f_\alpha(x) = \begin{cases} 4\alpha(x-1) - (1+\alpha+4\alpha x)\ln x, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

where $f_\alpha(x) = \frac{dF_\alpha(x)}{dx}$ and $-1 \le \alpha \le 1$. In Figure 4.2, the graphs of $F_\alpha(x)$ and $f_\alpha(x)$ are given for $x = 0.3, 0.5$, and 0.7.

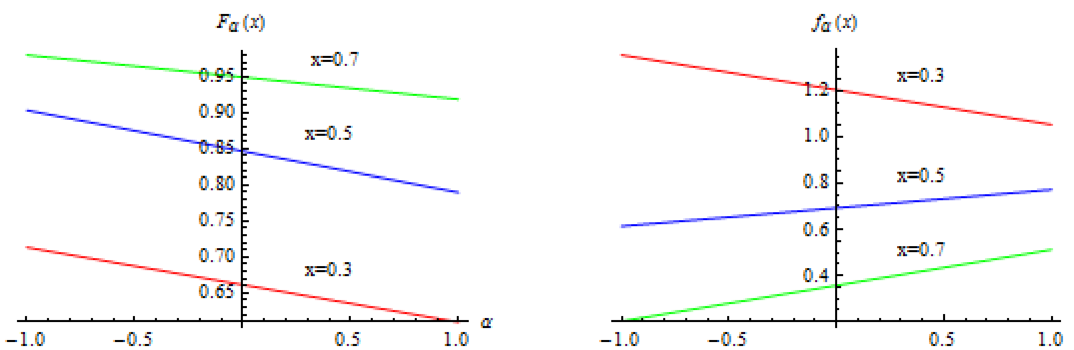

Figure 4.2: The graphs of $F_\alpha(x)$ and $f_\alpha(x)$ for selected values of x

These graphs show that larger values of α correspond to smaller values of $F_\alpha(x)$ for fixed x. The pdf $f_\alpha(x)$ varies depending on values of α such that for $0 < x < 0.42$, $f_\alpha(x)$ is decreasing and for $0.42 < x < 1$, $f_\alpha(x)$ is increasing with respect to α.

4.1.3 Asymptotic distributions of normalized exceedance statistics

In this section we discuss the asymptotic distribution of normalized exceedance statistics.

Let us define $S_m^* = \frac{S_m - E(S_m)}{\sqrt{Var(S_m)}}$. It is clear that $E(S_m^*) = 0$ and $Var(S_m^*) = 1$. Denote by $a_m = \frac{E(S_m)}{m}$, $b_m = \frac{\sqrt{Var(S_m)}}{m}$, $a = E(C_2(U,V))$, and $b = \sqrt{Var(C_2(U,V))}$. Then

$$\lim_{m\to\infty} \frac{E(S_m)}{\sqrt{Var(S_m)}} = \frac{a}{b}$$

and

$$\lim_{m\to\infty} \frac{m}{\sqrt{Var(S_m)}} = \frac{1}{b}.$$

Theorem 4.9 *The asymptotic distribution of the statistics S_m^* is given as*

$$\lim_{m\to\infty} \sup_{-\infty<x<\infty} \left| P\left\{S_m^* \le x\right\} - F_{a,b}^*(x) \right| = 0,$$

where

$$F_{a,b}^*(x) = \begin{cases} 0, & x < -\frac{a}{b} \\ P\left\{\frac{C_2(U,V)-a}{b} \le x\right\}, & x \in \left[-\frac{a}{b}, \frac{1-a}{b}\right] \\ 1, & x > \frac{a}{b} \end{cases}$$

Proof. Note that the probability $P\left\{\frac{S_m - E(S_m)}{\sqrt{Var(S_m)}} \le x\right\}$ can be written by

$$\begin{aligned} P\left\{\frac{S_m - E(S_m)}{\sqrt{Var(S_m)}} \le x\right\} &= P\left\{\frac{S_m/m - E(S_m)/m}{\sqrt{Var(S_m)}/m} \le x\right\} \\ &= P\left\{\frac{\frac{S_m}{m} - a_m}{b_m} \le x\right\} \end{aligned}$$

Since $0 \le a + bx \le 1$, it is obvious that

$$-\frac{a}{b} \le x \le \frac{1-a}{b}.$$

Hence

$$\begin{aligned} F_{a,b}^*(x) &\equiv \lim_{m\to\infty} P\left\{\frac{\frac{S_m}{m} - a_m}{b_m} \le x\right\} \\ &= P\left\{\frac{C_2(U,V) - a}{b} \le x\right\}, \ x \in \left[-\frac{a}{b}, \frac{1-a}{b}\right] \end{aligned}$$

and

$$\lim_{m\to\infty} \sup_{-\infty<x<\infty} |P\{S_m^* \le x\} - P\{C_2(U,V) \le a+bx\}| = 0.$$

□

Example 4.10 (Product Copula)**.** Let $C_1(u,v) = C_2(u,v) = uv$. Then

$$F_{a,b}^*(x) \equiv P\{C_2(U,V) \le a+bx\} = \iint\limits_{\{(u,v):uv\le a+bx\}} dudv.$$

Making the transformation $uv = t$ and $u = s$ with Jacobian $|J| = \frac{1}{s}$, we have

$$F_{a,b}^*(x) = \int_0^{a+bx} \int_t^1 \frac{1}{s} dsdt,$$

$$F_{a,b}^*(x) = \begin{cases} 0, & x < -\frac{a}{b} \\ -(a+bx)(\ln(a+bx)-1), & x \in \left[-\frac{a}{b}, \frac{1-a}{b}\right] \\ 1, & x > \frac{1-a}{b} \end{cases}$$

and

$$f_{a,b}^*(x) = \begin{cases} -b\ln(a+bx), & x \in \left[-\frac{a}{b}, \frac{1-a}{b}\right] \\ 0, & \text{otherwise} \end{cases}$$

Then

$$a = E(C_2(U,V)) = \int_0^1 \int_0^1 uvdudv = \frac{1}{4}$$

$$E(C_2^2(U,V)) = \int_0^1 \int_0^1 u^2v^2dudv = \frac{1}{9},$$
$$Var(C_2(U,V)) = \frac{7}{144},$$

and

$$b = \sqrt{Var(C_2(U,V))} = \frac{\sqrt{7}}{12}.$$

In Figure 4.3, the graphs of $F_{a,b}^*(x)$ and $f_{a,b}^*(x)$ in example 4.10 are illustrated.

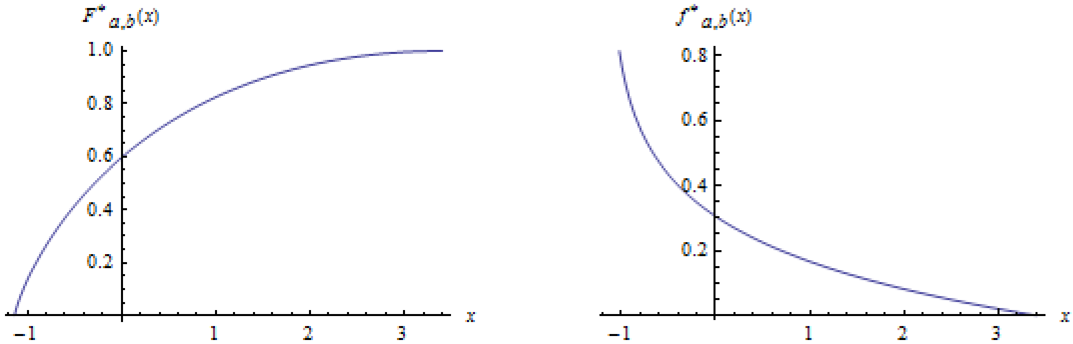

Figure 4.3: The graphs of $F^*_{a,b}(x)$ and $f^*_{a,b}(x)$ in example 4.10

Example 4.11 (FGM Copula). Let $C_1(u,v) = uv[1+\alpha(1-u)(1-v)]$ and $C_2(u,v) = uv$. Then

$$F^*_{a,b,\alpha}(x) = \lim_{m\to\infty} P\left\{S^*_m \le x\right\} = P\left\{C_2(U,V) \le a+bx\right\} = \iint\limits_{\{(u,v):uv\le a+bx\}} c_1(u,v)dudv.$$

$$= \iint\limits_{\{(u,v):uv\le a+bx\}} (1+\alpha(1-2u)(1-2v))dudv.$$

Making similar transformation as in example 4.10, yields

$$F^*_{a,b,\alpha}(x) = \int_0^{a+bx}\int_t^1 \left[1+\alpha(1-2s)\left(1-\frac{2t}{s}\right)\right]\frac{1}{s}dsdt$$

$$F^*_{a,b,\alpha}(x) = \begin{cases} 0, & x < -\frac{a}{b} \\ -(a+bx)(-3\alpha(a+bx-1) \\ +(1+\alpha+2a\alpha+2\alpha bx)\ln(a+bx)-1), & x \in \left[-\frac{a}{b}, \frac{1-a}{b}\right] \\ 1, & x > \frac{1-a}{b} \end{cases}$$

The limiting density function of S_m^* is

$$f_{a,b,\alpha}^*(x) = \begin{cases} b(4\alpha(a+bx-1) \\ -(1+\alpha+4a\alpha+4\alpha bx)\ln(a+bx), & x \in \left[-\frac{a}{b}, \frac{1-a}{b}\right] \\ 0, & \text{otherwise} \end{cases}.$$

Recall that we obtained from example 4.10 that, for $C_2(u,v) = uv$, $a = 0.25$ and $b = \frac{\sqrt{7}}{12} \approx 0.2205$. In Figure 4.4, the plots of $F_{a,b,\alpha}^*(x)$ are illustrated for selected values of x.

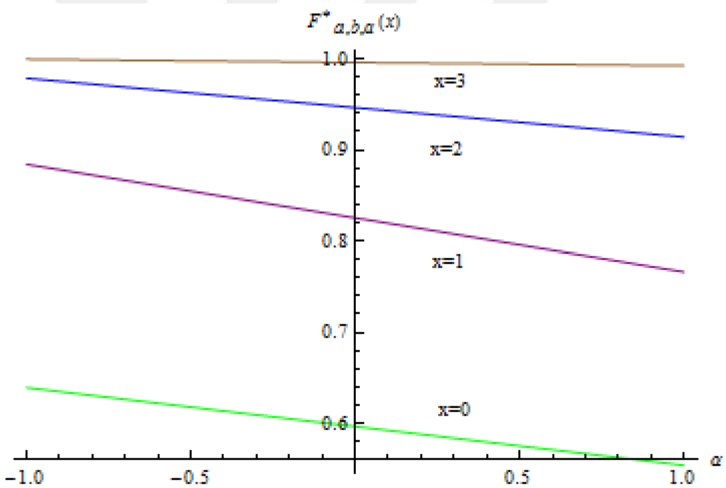

Figure 4.4: The graph of $F_{a,b,\alpha}^*(x)$ in example 4.11 for $a = 0.25$, $b = 0.2205$ and $-1 \le \alpha \le 1$.

It is observed that the function $F_{a,b,\alpha}^*(x)$ decreases, as α increases from -1 to 1.

4.2 Exceedance statistics based on order statistics and concomitants

In this section, we present the exact and asymptotic distributions of exceedance statistics based on order statistics and concomitants of bivariate sequences of random sequences.

4.2.1 Finite distributions of exceedance statistics based on order statistics and concomitants

Let $Z_1 = \{(X_i, Y_i),\ i = 1, 2, ..., n\}$ be a sequence of independent and identically distributed random vectors with joint distribution function $F(x, y) = C_1(F_X(x), F_Y(y))$. Similarly, let $Z_2 = \{(X_{n+j}, Y_{n+j}),\ j = 1, ..., m, ...\}$ be another sequence of independent and identically distributed random vectors with joint distribution function $G(x, y) = C_2(F_X(x),\ F_Y(y))$. Assume that the random vectors Z_1 and Z_2 are independent.

Let $(X_{r:n}, Y_{[r:n]})$ be the vector of the rth order statistic and its concomitant constructed from the sample Z_1 and define the binary random variables

$$\xi_i(r) \equiv I_N(X_i, Y_i) = \begin{cases} 1 & (X_{m+i}, Y_{m+i}) \in N, \\ 0 & \quad otherwise, \end{cases}$$

where $N \equiv (-\infty, X_{r:n}] \times (-\infty, Y_{[r:n]}]$.

Define the random variable $S_m(r) = \sum_{i=1}^{m} \xi_i(r)$ which is the number of observations of Z_2 falling into the set N. It is clear that the random variables $\xi_i(r)$, $i = 1, 2, ...$ are dependent. In the following theorem, we present the distribution of $S_m(r)$.

Theorem 4.10 *(See Erem and Bayramoglu [35]) The probability mass function of $S_m(r)$*

$$\begin{aligned} &P\{S_m(r) = k\} \\ &= \frac{1}{Beta(r, n-r+1)} \binom{m}{k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G(x,y)]^k [1 - G(x,y)]^{m-k} \\ &\times [F_X(x)]^{r-1} [1 - F_X(x)]^{n-r} dF_{X,Y}(x,y). \end{aligned} \tag{4.20}$$

The expression in terms of copulas is

$$\begin{aligned}
&P\left\{S_m(r)=k\right\}\\
&=\frac{1}{Beta\left(r,n-r+1\right)}\binom{m}{k}\int_0^1\int_0^1\left[C_2(u,v)\right]^k\left[1-C_2(u,v)\right]^{m-k}\\
&\times u^{r-1}[1-u]^{n-r}dC_1\left(u,v\right),
\end{aligned}\tag{4.21}$$

where $C_2(u,v)$ is a copula corresponding to joint distribution function $G(x,y)$ and $C_1(u,v)$ is a copula corresponding to joint distribution function $F(x,y)$.

Proof. The proof of the theorem is very similar to Theorem 4.3. The only difference is that here we use the joint pdf of $(X_{r:n},Y_{[r:n]})$ instead of the joint pdf of (X,Y). First we show that

$$\begin{aligned}
&P\left\{S_m(r)=k\right\}\\
&=\frac{1}{Beta\left(r,n-r+1\right)}\binom{m}{k}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[G(x,y)\right]^k\left[1-G(x,y)\right]^{m-k}\\
&\times[F_X(x)]^{r-1}[1-F_X(x)]^{n-r}dF_{X,Y}\left(x,y\right)
\end{aligned}$$

Define

$$A_{i_j}=\left\{X_{i_j}<X_{r:n},Y_{i_j}<Y_{[r:n]}\right\}$$

and observe that the complement of A_{i_j} as

$$A_{i_j}^c=\left\{X_{i_j}<X_{r:n},Y_{i_j}>Y_{[r:n]}\right\}\cup\left\{X_{i_j}>X_{r:n},Y_{i_j}<Y_{[r:n]}\right\}\cup\left\{X_{i_j}>X_{r:n},Y_{i_j}>Y_{[r:n]}\right\}.$$

For simplicity, we define the random set E as

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

Then conditioning with respect to $X_{r:n} = x$ and $Y_{[r:n]} = y$ we obtain

$$\begin{aligned}P\{S_m(r) = k\} &= \sum_{i_1,i_2,...,i_m} P\left\{A_{i_1}A_{i_2}...A_{i_k}A^c_{i_{k+1}}...A^c_{i_m}\right\} \\ &= \sum_{i_1,i_2,...,i_m} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{E \mid X_{r:n} = x, Y_{[r:n]} = y\right\} dF_{X_{r:n},Y_{[r:n]}}(x, y) \\ &= \sum_{i_1,i_2,...,i_m} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{a_{i_1}a_{i_2}...a_{i_k}a^c_{i_{k+1}}...a^c_{i_m}\right\} dF_{X_{r:n},Y_{[r:n]}}(x, y),\end{aligned} \tag{4.22}$$

where the sum $\sum_{i_1,i_2,...,i_m}$ extends over all permutations of $i_1, i_2, ..., i_m \in \{1, 2, ..., m\}$, and

$$a_{i_j} = \left\{X_{i_j} < x, Y_{i_j} < y\right\}, \; a^c_{i_j} \text{ is the complement of } a_{i_j}.$$

Since events a_{i_j} are independent the probability under integral in (4.22) can be written as

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

Using the joint probability density function of $X_{r:n}$ and $Y_{[r:n]}$ in equation (4.22), we obtain

$$\begin{aligned}&P\{S_m(r) = k\} \\ &= \frac{1}{Beta\,(r, n-r+1)}\binom{m}{k}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[G(x, y)]^k\,[1 - G(x, y)]^{m-k} \\ &\times f\,(y \mid x)\, f_{X_{r:n}}\,(x)\,dxdy.\end{aligned}$$

Therefore, the finite sample distribution of $S_m(r)$ can be simplified as

$$\begin{aligned}&P\{S_m(r) = k\} \\ &= \frac{1}{Beta\,(r, n-r+1)}\binom{m}{k}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[G(x, y)]^k\,[1 - G(x, y)]^{m-k} \\ &\times [F_X(x)]^{r-1}[1 - F_X(x)]^{n-r}dF_{X,Y}\,(x, y)\,.\end{aligned}$$

Using the probability integral transformation $F_X(x) = u$ and $F_Y(y) = v$ in equation (4.20), we obtain equation (4.21).

□

Proposition 4.11 *It is true that*

$$E(S_m(r)) = mE\left(G(X_{r:n}, Y_{[r:n]})\right) = mE(C_2(U_{r:n}, V_{r:n})), \tag{4.23}$$

$$\begin{aligned} Var(S_m(r)) &= mE(G(X_{r:n}, Y_{[r:n]})) - mE(G^2(X_{r:n}, Y_{[r:n]})) + m^2 Var(G(X_{r:n}, Y_{[r:n]})) \\ &= mE(C_2(U_{r:n}, V_{[r:n]})) - mE(C_2^2(U_{r:n}, V_{[r:n]})) + m^2 Var(C_2(U_{r:n}, V_{[r:n]})), \end{aligned} \tag{4.24}$$

where $U_{r:n}$ and $V_{[r:n]}$ are respectively the rth order statistic and its concomitant constructed from the random sample (U_1, V_1),(U_2, V_2),...,(U_n, V_n) with distribution $C_2(u, v)$.

Proof. One can write equation (4.20) in the following form

$$P\{S_m(r) = k\} = \binom{m}{k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G(x, y)]^k [1 - G(x, y)]^{m-k} dF_{X_{r:n}, Y_{[r:n]}}(x, y).$$

Using this probability function, the expected value of $S_m(r)$ can be calculated directly as

$$\begin{aligned} E(S_m(r)) &= \sum_{k=0}^{m} kP\{S_m(r) = k\} \\ &= \sum_{k=0}^{m} k\binom{m}{k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G(x, y)]^k [1 - G(x, y)]^{m-k} dF_{X_{r:n}, Y_{[r:n]}}(x, y). \\ &= \sum_{k=0}^{m} k\binom{m}{k} E\left[[G(X_{r:n}, Y_{[r:n]})]^k \left[1 - G(X_{r:n}, Y_{[r:n]})\right]^{m-k}\right] \\ &= E\left[\sum_{k=0}^{m} k\binom{m}{k} [G(X_{r:n}, Y_{[r:n]})]^k \left[1 - G(X_{r:n}, Y_{[r:n]})\right]^{m-k}\right] \\ &= mE\left(G(X_{r:n}, Y_{[r:n]})\right). \end{aligned} \tag{4.25}$$

Similarly, the second moment of $S_m(r)$ is

$$\begin{aligned}
E(S_m^2(r)) &= \sum_{k=0}^{m} k^2 P\left\{S_m(r) = k\right\} \\
&= \sum_{k=0}^{m} k^2 \binom{m}{k} E\left[\left[G(X_{r:n}, Y_{[r:n]})\right]^k \left[1 - G(X_{r:n}, Y_{[r:n]})\right]^{m-k}\right] \\
&= E\left[\sum_{k=0}^{m} k^2 \binom{m}{k} \left[G(X_{r:n}, Y_{[r:n]})^k\right] \left[1 - G(X_{r:n}, Y_{[r:n]})\right]^{m-k}\right] \\
&= m^2 E(G^2(X_{r:n}, Y_{[r:n]})) + mE\left[G(X_{r:n}, Y_{[r:n]})(1 - G(X_{r:n}, Y_{[r:n]}))\right].
\end{aligned}$$

Therefore using first two moments, we can calculate the variance of $S_m(r)$.

$$\begin{aligned}
Var\left(S_m(r)\right) &= m^2 \left[E\left(G(X_{r:n}, Y_{[r:n]})\right)\right]^2 - m^2 E(G^2(X_{r:n}, Y_{[r:n]})) \\
&\quad - mE\left[G(X_{r:n}, Y_{[r:n]})(1 - G(X_{r:n}, Y_{[r:n]}))\right] \\
&= Var(G(X_{r:n}, Y_{[r:n]})) - mE\left(G(X_{r:n}, Y_{[r:n]})\right) - mE(G^2(X_{r:n}, Y_{[r:n]})).
\end{aligned} \tag{4.26}$$

Using the probability integral transformation in equations (4.25) and (4.26), the proof is completed. □

Corollary 4.12 *If $C_1(u,v) = C_2(u,v) = C(u,v)$, then*

$$P\left\{S_m(r) = k\right\}$$
$$= \frac{1}{Beta\left(r, n-r+1\right)} \binom{m}{k} \int_0^1 \int_0^1 \left[C(u,v)\right]^k \left[1 - C(u,v)\right]^{m-k} u^{r-1}[1-u]^{n-r} dC(u,v).$$

4.2.1.1 Numerical Results.

Below in Table 4.2 numerical results for the pmf of the exceedance statistics $S_m(r)$ for different copulas and for $m = n = 5$ and $r = 3$ are presented. These copulas are given in the following:

Case i): $C_1(u,v) = C_2(u,v) = uv$.

Case ii): $C_1(u,v) = uv$, $C_2(uv) = uv\exp(-\theta \ln u \ln v), \theta \in (0,1]$ (Gumbel-Barnett copula).

Case iii): $C_1(u,v) = uv, C_2(u,v) = [\max(u^{-\theta}+v^{-\theta}-1,0)]^{-1/\theta}, \theta \in [-1,\infty)\backslash\{0\}$ (Clayton copula).

Case iv): $C_1(u,v) = uv$, $C_2(u,v) = \frac{uv}{1-\theta(1-u)(1-v)}$, $\theta \in [-1,1)$ (Ali-Mikhail-Haq copula).

	Case i	Case ii θ=0.5	Case iii θ=1	Case iv θ=-1
	$P\{S_5(3)=k\}$	$P\{S_5(3)=k\}$	$P\{S_5(3)=k\}$	$P\{S_5(3)=k\}$
k = 0	0.361	0.036	0.268	0.4175
k = 1	0.278	0.065	0.284	0.362
k = 2	0.188	0.104	0.226	0.168
k = 3	0.109	0.161	0.140	0.0451
k = 4	0.050	0.247	0.064	0.0070
k = 5	0.014	0.387	0.018	0.0004

Table 4.2: Numerical results for the exact distribution of $S_m(r)$

As it is seen from Table 4.2, while k increases, pmf of S_m decreases in cases i) and iv). However, pmf of S_m increases in case ii), as k increases.

4.2.2 Asymptotic distributions of exceedance statistics based on order statistics and concomitants

In this section, we derive the asymptotic distribution of $\frac{S_m(r)}{m}$ as $m \to \infty$. The following theorem presents a result for $\lim_{m\to\infty} P\left\{\frac{S_m(r)}{m} \leq x\right\}$.

Theorem 4.13 *(See Erem and Bayramoglu [35]) The statistics* $\frac{S_m(r)}{m}$ *has the continuous limiting distribution*

$$T(x)=\begin{cases} 0, & x<0 \\ P\left\{C_2(U_{r:n},V_{r:n})\leq x\right\}, & 0\leq x\leq 1 \\ 1, & x>1 \end{cases}, \tag{4.27}$$

where $U_{r:n}$ *and* $V_{[r:n]}$ *are respectively the* r*th order statistic and its concomitant constructed from the random sample* (U_1,V_1),(U_2,V_2),...,(U_n,V_n) *with distribution* $C_2(u,v)$.

The expression in terms of joint cdf is

$$T(x)=\begin{cases} 0, & x<0 \\ P\left\{G(X_{r:n},Y_{[r:n]})\leq x\right\}, & 0\leq x\leq 1 \\ 1, & x>1 \end{cases}.$$

Proof. The proof of the theorem is similar to Theorem 4.10. The only difference that here we use the joint pdf of $(X_{r:n},Y_{[r:n]})$ instead of the joint pdf of (X,Y). $\square$

Corollary 4.14 *If* $C_1(u,v)=C_2(u,v)=C(u,v)$ *then*

$$T(x)=\begin{cases} 0, & x<0 \\ P\left\{\frac{1}{Beta(r,n-r+1)}\iint\limits_{\{(t,s):\ C(t,s)\leq x\}} c\,(t,s)\,t^{r-1}\,(1-t)^{n-r}\,dtds\leq x\right\}, & 0\leq x\leq 1 \\ 1, & x>1 \end{cases},$$

where

$$c(t,s)=\frac{\partial^2 C(t,s)}{\partial t\partial s}.$$

Proof. Since $(X_{r:n}, Y_{[r:n]})$ is a vector of the rth order statistics and its concomitant constructed from the sample Z_1 with joint distribution function $F(x, y) = C(F_X(x), F_Y(y))$, we can use the joint pdf of $X_{r:n}$ and $Y_{[r:n]}$ given in equation (2.11). Consequently,

$$\begin{aligned} P\left\{G\left(X_{r:n}, Y_{[r:n]}\right) \leq x\right\} &= \iint\limits_{\{(u,v):\ G(u,v)\leq x\}} f\left(v \mid u\right) f_{r:n}\left(u\right) dudv \\ &= \frac{1}{Beta\left(r, n-r+1\right)} \iint\limits_{\{(u,v):\ G(u,v)\leq x\}} \frac{f_{X,Y}\left(u, v\right)}{f_X\left(u\right)} \\ &\times F_X^{r-1}\left(u\right)\left[1 - F_X\left(u\right)\right]^{n-r} dF_X\left(u\right) dv \end{aligned} \tag{4.28}$$

and the result follows. □

Example 4.12 (Product copula)**.** Let $C(u, v)$ be the product copula, i.e. the random variables X_i and Y_i are independent. Then from equation (4.28) we obtain

$$\begin{aligned} T_1(x) &\equiv \lim_{m\to\infty} P\left\{\frac{S_m(r)}{m} \leq x\right\} \\ &= \frac{1}{Beta\left(r, n-r+1\right)} \iint\limits_{\{(u,v):\ uv\leq x\}} u^{r-1}\left(1-u\right)^{n-r} dudv \end{aligned}$$

Making transformation $t = u$ and $s = uv$ with Jacobian $|J| = \frac{1}{t}$, we have

$$T_1(x) = \frac{1}{Beta\left(r, n-r+1\right)} \int_0^x \int_s^1 t^{r-2}\left(1-t\right)^{n-r} dtds. \tag{4.29}$$

For example if $r = 3$ and $n = 5$, the limiting distribution of $\frac{S_m(r)}{m}$ is

$$T_1(x) = \frac{x\left(-3x^4 + 10x^3 - 10x^2 + 5\right)}{2},\ 0 \leq x \leq 1.$$

Example 4.13 (Clayton copula)**.** Let $C(u,v)$ be a Clayton copula, i.e.,

$$C\left(u,v\right)=\left(u^{-1/c}+v^{-1/c}-1\right)^{-c},\quad c>0.$$

For $c=1$, the clayton copula becomes

$$C\left(u,v\right)=\frac{uv}{u+v-uv},$$

and

$$c\left(u,v\right)=-\frac{2uv}{\left(u\left(v-1\right)-v\right)^{3}}.$$

Then for $r=3$ and $n=5$, from equation (4.28) we obtain

$$\begin{aligned}
&T_2(x)\equiv\lim_{m\to\infty}P\left\{\frac{S_m(r)}{m}\leq x\right\}\\
&=\frac{1}{Beta\left(3,3\right)}\int_0^x\int_0^{1-t}2ts^2dsdt\\
&=x^2\left(-4x^3+15x^2-20x+10\right),\ 0\leq x\leq 1.
\end{aligned}$$

Example 4.14 (Ali-Mikhail-Haq copula)**.** (Nelson, 2006, p. 116) Let $C(u,v)$ be an Ali-Mikhail-Haq copula, i.e.,

$$C\left(u,v\right)=\frac{uv}{1-\alpha\left(1-u\right)\left(1-v\right)},\quad -1\leq\alpha<1.$$

Let $\alpha=-1$, $r=3$ and $n=5$. From equation (4.28) the limiting distribution of

$\frac{S_m(r)}{m}$ can be calculated for all x, $0 \leq x \leq 1$ as

$$\begin{aligned} T_3(x) &\equiv \lim_{m\to\infty} P\left\{\frac{S_m(r)}{m} \leq x\right\} \\ &= \frac{1}{Beta\,(3,3)} \int_0^{1-t} \int_0^{x} \frac{2\,(1+st-s-t)\,s^2}{1+s} dtds \\ &= x(-4x^4 + 5x^3 - 20x^2 + 100x \\ &+ 60(2-x)\ln(2-x) - 80). \end{aligned}$$

In Table 4.3, the limiting distributions of $P\left\{\frac{S_m(r)}{m} \leq x\right\}$ for selected copulas are presented:

C(u,v)	n	r	$\lim_{m\to\infty} P\left\{\frac{S_m(r)}{m} \leq x\right\}$
uv	5	3	$T_1(x) = \frac{x\left(-3x^4+10x^3-10x^2+5\right)}{2}$
$(u^{-1}+v^{-1}-1)^{-1}$	5	3	$T_2(x) = x^2(-4x^3+15x^2-20x+10)$
$\frac{uv}{1+(1-u)(1-v)}$	5	3	$T_3(x) = x(-4x^4+5x^3-20x^2+100x+60(2-x)\ln(2-x)-80)$

Table 4.3: Asymptotic distributions of $\frac{S_m(r)}{m}$ for product, Clayton and Ali-Mikhail-Haq copulas

$T_1(x)$, $T_2(x)$ and $T_3(x)$ are new polynomial continuous distributions arising in considered exceedance models. In Figure 4.5 the graphical representations of these functions are illustrated below.

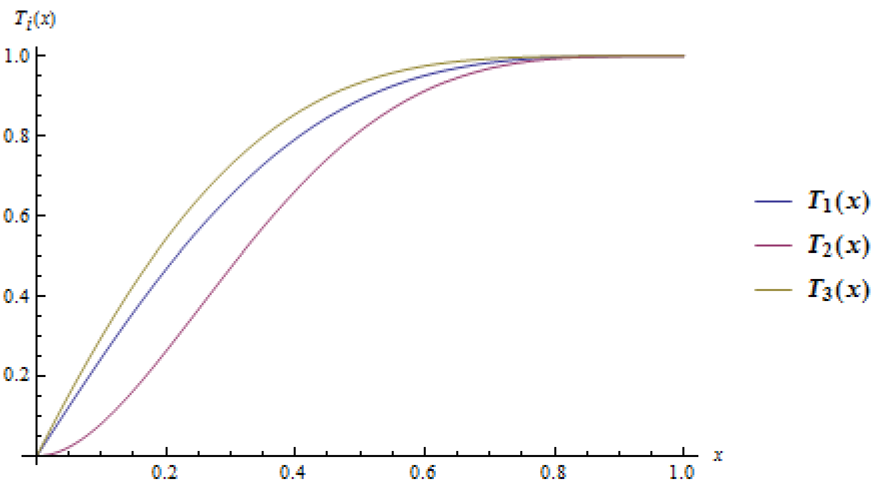

Figure 4.5: The graphics of limiting distributions given in Table 4.3

In Figure 4.5, the $T_1(x)$, $T_2(x), T_3(x)$ are the limiting distributions of $\frac{S_m(r)}{m}$ with underlying copulas $C(u,v) = uv$, $C(u,v) = (u^{-1}+v^{-1}-1)^{-1}$, and $C(u,v) = \frac{uv}{1+(1-u)(1-v)}$, respectively, for $n = 5$ and $r = 3$.

4.2.3 Asymptotic distributions of normalized exceedance statistics based on order statistics and concomitants

Now, as in Section 4.1.3, asymptotic distribution of normalized exceedance statistics are investigated.

Let us define $S_m^*(r) = \frac{S_m(r)-E(S_m(r))}{\sqrt{Var(S_m(r))}}$. It is clear that $E(S_m^*(r)) = 0$ and $Var(S_m^*(r)) = 1$. Denote by $a_m(r) = \frac{E(S_m(r))}{m}$, $b_m(r) = \frac{\sqrt{Var(S_m(r))}}{m}$. Also denote by, $a(r) = E(C_2(U_{r:n}, V_{[r:n]}))$ and $b(r) = \sqrt{Var(C_2(U_{r:n}, V_{[r:n]}))}$. It is clear that

$$\lim_{m\to\infty} \frac{E(S_m(r))}{\sqrt{Var(S_m(r))}} = \frac{a(r)}{b(r)}$$

and

$$\lim_{m\to\infty} \frac{m}{\sqrt{Var(S_m(r))}} = \frac{1}{b(r)}.$$

Theorem 4.15 *The statistics $S_m^*(r)$ has the continuous limiting distribution function*

$$H(x) = \begin{cases} 0, & x < -\frac{a(r)}{b(r)} \\ P\left\{\frac{C_2(U_{r:n}, V_{[r:n]}) - a(r)}{b(r)} \leq x\right\}, & x \in \left[-\frac{a(r)}{b(r)}, \frac{1-a(r)}{b(r)}\right] \\ 1, & x > \frac{1-a(r)}{b(r)} \end{cases}$$

Proof. Proof of this Theorem is similar to the proof of Theorem 4.9. □

Example 4.15 (Product copula). Let $C_1(u,v) = C_2(u,v) = uv$. Then the continuous limiting distribution function of $S_m^*(r)$ is given by

$$H(x) = \frac{1}{Beta(r, n-r+1)} \int_0^{a(r)+b(r)x} \int_t^1 s^{r-2}(1-s)^{n-r} ds dt.$$

The mean of $C_2\left(U_{r:n}, V_{[r:n]}\right)$ is

$$
\begin{aligned}
E\left(C_2(U_{r:n}, V_{[r:n]})\right) &= \int_0^1 \int_0^1 C_2(u,v) dF_{U_{r:n},V_{[r:n]}}(u,v) \\
&= \int_0^1 \int_0^1 uv f_{U_{r:n},V_{[r:n]}}(u,v) dudv \\
&= \int_0^1 \int_0^1 uvf\left(v \mid u\right) f_{r:n}\left(u\right) dudv \\
&= \frac{1}{Beta\left(r, n-r+1\right)} \int_0^1 \int_0^1 vu^r (1-u)^{n-r} dudv \\
&= \frac{1}{Beta\left(r, n-r+1\right)} \int_0^1 u^r (1-u)^{n-r} du \qquad (4.30) \\
&= \frac{1}{Beta\left(r, n-r+1\right)} Beta(r+1, n-r+1) \qquad (4.31) \\
&= \frac{r}{2(n+1)}. \qquad (4.32)
\end{aligned}
$$

and the second moment of $C_2(U_{r:n}, V_{[r:n]})$ is calculated as

$$
\begin{aligned}
E\left(C_2^2(U_{r:n}, V_{[r:n]})\right) &= \int_0^1 \int_0^1 C_2^2(u,v) dF_{U_{r:n},V_{[r:n]}}(u,v) \\
&= \frac{1}{Beta\left(r, n-r+1\right)} \int_0^1 \int_0^1 v^2 u^{r+1} (1-u)^{n-r} dudv \\
&= \frac{1}{3Beta\left(r, n-r+1\right)} \int_0^1 u^{r+1} (1-u)^{n-r} du \\
&= \frac{1}{3Beta\left(r, n-r+1\right)} Beta(r+2, n-r+1) \\
&= \frac{n!}{3(r-1)!(n-r)!} \frac{(r+1)!(n-r)!}{(n+2)!} \\
&= \frac{r(r+1)}{3(n+1)(n+2)}. \qquad (4.33)
\end{aligned}
$$

Consequently, using the first two moments of $C_2(U_{r:n}, V_{[r:n]})$ we can calculate the variance of $C_2(U_{r:n}, V_{[r:n]})$ as

$$\begin{aligned}
Var\left(C_2(U_{r:n}, V_{[r:n]})\right) &= E\left(C_2^2(U_{r:n}, V_{[r:n]})\right) - \left[E\left(C_2(U_{r:n}, V_{[r:n]})\right)\right]^2 \\
&= \frac{r(r+1)}{3(n+1)(n+2)} - \left(\frac{r}{2(n+1)}\right)^2 \\
&= \frac{4r(r+1)(n+1) - 3r^2(n+2)}{12(n+1)^2(n+2)} \\
&= \frac{r^2n - 2r^2 + 4rn + 4r}{12(n+1)(n+2).} \qquad (4.34)
\end{aligned}$$

Analytical expression of function $H(x)$ is given in integral form. We present below, in Figure 4.6, the graphs of $H(x)$ for $n = 4$ and $r = 1, ..., 4$.

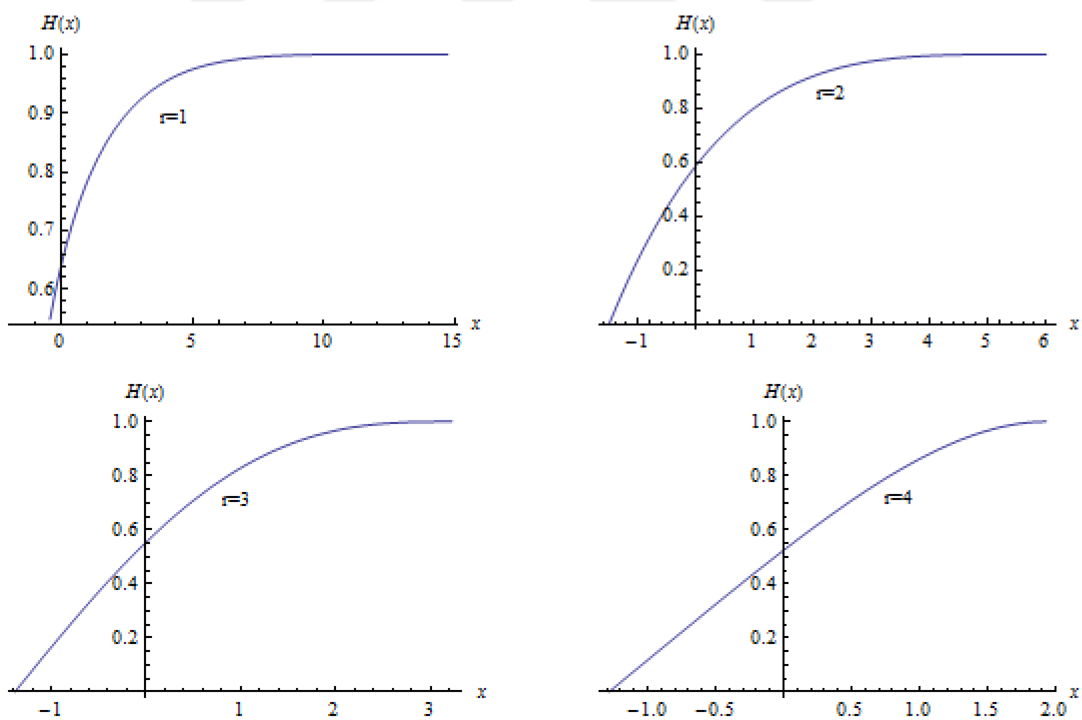

Figure 4.6: The graphs of $H(x)$ for $C_1(u, v) = C_2(u, v) = uv$ for $n = 4$ and $r = 1, ..., 4$

Chapter 5

Application

Hypertension has long been a serious health problem, as one of the most important causes of cardiovascular heart disease, and has an adverse effects on life quality. Therefore, developing countries have taken precautions for preventing hypertension, including promoting healthy nutrition, physical exercise, avoiding stress. The classification of blood pressure is given in the following table:

	Systolic Blood Pressure (SBP)	Diastolic Blood Pressure (DBP)
Normal	$\leq 120\ mm\ Hg$	$\leq 80\ mm\ Hg$
Prehypertension	$120 - 139\ mm\ Hg$	$80 - 89\ mm\ Hg$
Hypertension	$\geq 140\ mm\ Hg$	$\geq 90\ mm\ Hg$

Table 5.1: Classification of blood pressure levels

The values are given in Table 5.1 are the thresholds for hypertension (see Urden et al., 2014; p. 341). Let the sequence of random vectors $Z_1 = \{(X_i, Y_i)\,,$ $i = 1, 2, ..., n\}$ be a training sample consisting of hypertensive patients and interpret X_i and Y_i as SBP and DBP of ith patient. Assume that $X \sim U(120, 260)$ and $Y \sim U(90, 150)$. Then, consider another sequence of random vectors $Z_2 = \{(X_{n+j}, Y_{n+j}),\ j = 1, 2, ..., m, ...\}$ as a control sample, according to which, patients will be categorized as hypertensive or not. Therefore, the statistics S_m and $S_m(r)$ counts the total number of people with normal blood pressure in control sample. Also, the proposed exceedance statistics can be used to make inferences about prevalence of hypertension. In this way, the number of hypertensive people can be accurately estimated.

Recently, air pollution has become an important problem, causing habitat loss and respiratory illness. Especially, particular matter (PM) pollution has a negative effect on the ecological balance. The sum of liquid and solid particles of varying sizes are called as particular matter (PM). Particular matter is classified as coarse dust (PM_{10}) (2.5 to 10 micrometers), and fine particles (less than 2.5 micrometers) ($PM_{2.5}$).

In 2006 World Health Organization (WHO) determined the threshold values for $PM_{2.5}$ and PM_{10} (see WHO Air Quality Guidelines [68]). While the average permitted value for $PM_{2.5}$ is $25mg/m^3$, it is $50mg/m^3$ for PM_{10}. According to [68], the average values of PM_{10} should not surpass the determined thresholds for more than 35 days in a year. Assume that daily levels of $PM_{2.5}$ and PM_{10} are independent from each other days. Let $Z = \{(X_i, Y_i), i = 1, 2, ..., m, ...\}$ be iid bivariate observations. Then, X_i and Y_i can be interpreted as the daily average values of $PM_{2.5}$ and PM_{10}, respectively. Consider,

$$\xi_i = \begin{cases} 1 & if\ (X_i, Y_i) \in (-\infty, X) \times (-\infty, Y), \\ 0 & \text{otherwise.} \end{cases}$$

Consequently, $S_m = \sum_{i=1}^{m} \xi_i$ is the total number of days in which average values of $PM_{2.5}$ and PM_{10} does not surpass the critical threshold values.

Chapter 6

Conclusion

Random threshold models provide a basis for hypothesis testing and statistical inference. They also play an important role in real life problems such as reliability, modelling hydrological events and air pollution. In the literature, there have been many studies about univariate random threshold models. However, in modelling real life problems, there is a need for multivariate random threshold models.

In this thesis bivariate random threshold models based on bivariate random sequences are investigated. Later, the finite and asymptotic distributions of exceedance statistics are studied. Because of considered bivariate random variables, the obtained distributions are expected to depend on copulas. These result agree with the models in real life applications.

The discussion of the use of these proposed models in medicine and air pollution illuminate the application area of exceedance statistics and random threshold models.

BIBLIOGRAPHY

[1] Arnold, B.C., N. Balakrishnan and H. N. Nagaraja. 2008. *A First Course in Order Statistics (Classics in Applied Mathematics)*. USA: SIAM.

[2] Ahsanullah, M. 1979. *Characterization of Exponential Distribution by Record Values*. Sankhya, Vol. 41 B: 116–121.

[3] Ahsanullah, M. 1991. *Record Values of Lomax Distribution*. Statistica Neerlandica. Vol. 45(1): 21–29.

[4] Ahsanullah, M. 1992. *Record Values of Independent and Identically Distributed Continuous Random Variables*. Pak. J. Statist., Vol. 8(2) A: 9–34.

[5] Ahsanullah, M. 1995. *Record Statistics*. USA: Nova Science Publishers.

[6] Ahsanullah, M. and V.B. Nevzorov. 2015. *Records via Probability Theory*. Amsterdam: Atlantiss Press.

[7] Aitken, A.C. 1939. Determinants and Matrices. Edinburgh: Oliver & Boyd.

[8] Bairamov, I.G. and Yu.I. Petunin. 1991. *Structure of Invariant Confidence Intervals Containing the Main Distributed Mass*. Theory of Probability and its Applications, Vol. 35(1): 15–26.

[9] Bairamov, I.G. 1997. *Some Distribution Free Properties of Statistics Based on Record Values and Characterizations of the Distributions Through a Record*. Journal of Applied Statistical Science, Vol. 5(1): 17–25.

[10] Bairamov, I.G., O.L. Gebizlioglu, and M.F. Kaya. 1999. *Distributional Properties of Statistics Based on Invariant Confidence Intervals*. İstatistik, Journal of the Turkish Statistical Association, Vol. 2(1): 15-32.

[11] Bairamov I.G. and S. Kotz 2001. *On Distributions of Exceedances Associated with Order Statistics and Record Values for Arbitrary Distributions*. Statistical Papers, Vol. 42: 171–185.

[12] Bairamov, I. and S. Eryılmaz. 2004. *Characterization of Symmetry and Exceedance Models in Multivariate FGM Distributions*. Journal of Applied Statistical Science, Vol. 13(2): 87–99.

[13] Bairamov, I.G. and M. K. Khan. 2007. *On Exceedances of Record and Order Statistics*, Proceedings of the American Mathematical Society, Vol. 135(6): 1935–1945.

[14] Bairamov, I. and H. Tanil. 2007. *On Distributions of Exceedances Based on Generalized Order Statistics*. Communications in Statistics - Theory and Methods, Vol. 36(7): 1479–1491.

[15] Bairamov, I. G. and S. Eryılmaz. 2008. *Joint Behaviour of Precedences in a Random Threshold Model*. Australian and New Zealand Journal of Statistics, Vol. 50(3): 209–216.

[16] Bairamov, I. and S. Eryılmaz. 2009. *Waiting Times of Exceedances in Random Threshold Models*. Statistics and Probability Letters, Vol. 79(5): 676–683.

[17] Balakrishnan, N. and C.R. Rao. 1998. Handbook of Statistics 16: Order Statistics-Theory and Methods. Technometrics, Newyork: Elsevier.

[18] Balakrishnan, N. and H.K. Tony Ng. 2006. *Precedence-Type Tests and Applications*. USA: John Wiley & Sons.

[19] Balakrishnan, N. 2007. *Permanent, Order Statistics, Outliers and Robustness, Review*. Mathematics of Computation, Vol. 20(1): 7–107.

[20] Balakrishnan, N. and C.D. Lai. 2009. *Continuous Bivariate Distributions (2^{nd} ed.)*. Newyork: Springer.

[21] Balakrishnan, N. and V.B. Nevzorov. 1997. *Advances in Combinatorial Methods and Applications to Probability and Statistics (N. Balakrishnan, ed.):* Statistics and Technology Series, pp. 189–200, Boston: Birkhauser.

[22] Bayramoglu, I. and G. Giner. 2014. *Order Statistics and Exceedances for Some Models of INID Random Variables*. Brazilian Journal of Probability and Statistics, Vol. 28(4): 492–514.

[23] Benestad, R. E. 2003. *How Often Can We Expect a Record Event?*. Climate Research, Vol. 25(1): 3–13.

[24] Benestad, R. E. 2004. *Record-values, Nonstationarity Tests and Extreme Value Distributions*. Global and Planetary Change, Vol. 44 (1–4): 11–26.

[25] Benestad R. E. 2006. *Can We Expect More Extreme Precipitation on the Monthly Time Scale?*. Journal of Climate, Vol: 19: 630–637.

[26] Borovkov, A.A. 1998. *Mathematical Statistics*. Amsterdam: Gordon and Breach Science Publishers.

[27] Boutsikas, M.V. and M.V. Koutras. 2002. *Modeling Claim Exceedances over Thresholds*. Insurance: Mathematics and Economics, Vol. 30(1): 67–83.

[28] Casella, G. and R.L. Berger. 2002. *Statistical Inference* (2^{nd} edition). CA, Belmond: Duxbury Press.

[29] Cherubini, U. vd. 2012. *Dynamic Copula Methods in Finance*. India: Wiley.

[30] Danaher, P.J. and M.S. Smith. 2010. *Modeling Multivariate Distributions Using Copulas: Applications in Marketing*. Marketting Science, Vol. 30(1): 4–21.

[31] David, H.A., M.J. O' Connell, and S.S. Yang. 1977. *Distribution and Expected Value of the Rank of a Concomitant of an Order Statistic*. Ann Statist. 5(1): 216–223.

[32] David, H.A. and H.N. Nagaraja. 1998. *Concomitans of Order Statistics, In Handbook of Statistics 16: Order Statistics: Theory & Methods (Eds. N. Balakrishnan and C.R. Rao)*. Amsterdam: North Holland Academic Publishers.

[33] David, H.A. and H.N. Nagaraja. 2003. *Order Statistics* (3^{rd} edition). New Jersey: John Wiley & Sons.

[34] Epstein, B. 1954. *Tables for the Distribution of the Number of Exceedances.* Ann. Math. Statist., Vol. 25(4): 762–768.

[35] Erem, A. and I. Bayramoglu. (2017). *Exact and Asymptotic Distributions of Exceedance Statistics for Bivariate Random Sequences*. Statistics and Probability Letters, Vol. 125: 181–188.

[36] Eryılmaz, S. and I.G. Bairamov. 2003. *On a New Sample Rank of an Order Statistics and Its Concomitant*. Statistics & Probability Letters, Vol. 63(2): 123–131.

[37] Eryilmaz, S., O.L. Gebizlioglu, and F. Tank. 2011. *Modelling of Claim of Exceedances over Random Thresholds for Related Insurance Portfolios*. Insurance: Mathematics and Economics 49(3): 496–500.

[38] Genest, C. and L.P. Rivest. 1993. *Statistical Inference Procedures for Bivariate Archimedean Copulas*. Journal of the American Statistical Association, Vol. 88(423): 1034–1043.

[39] Gumbel, E.J. 1954a. *Statistical Theory of Extreme Values and Some Practical Applications*. Washington: National Bureau of Standards.

[40] Gumbel, E.J. 1954b. *Elementare Ableitung der Momente für die Zahl der Überschreitungen*. Mitt. Math. Statist., Vol. 6: 164–169.

[41] Gumbel, E.J. and H. Von Schelling. 1950. *The Distribution of the Number of Exceedances*. Ann. Math. Statist. Papers, Vol. 21(3): 247–262.

[42] He, Q. and H.N. Nagaraja. 2009. *Distribution of Concomitants and Their Order Statistics*. Journal of Statistical Planning and Inference, Vol. 139(8): 2643–2655.

[43] Johnson, N. and S. Kotz. 1991. *Some Generalizations of Bernoulli and Polya-Eggenberger Contagion Models*. Statist. Papers, Vol. 32(1): 1–17.

[44] Johnson, N. and S. Kotz. 1994. *Further Comments on Matveychuk and Petunin' s Generalized Bernoulli model, and Nonparametric Tests of Homogenity*. J. Statist. Plann. Inference, Vol. 41(1): 61–72.

[45] Katzenbeisser, W. 1985. *The Distribution of Two-Sample Location Exceedance Test Statistics Under Lehmann alternatives*. Statistische Hefte, Vol. 26(1): 131-138.

[46] Katzenbeisser, W. 1986. *The Exact Power of Two-Sample Location Test Based on Exceedance Against Shift Alternatives*. Math. Op. Forsch (Statistics), 20(1): 47–54.

[47] Kemalbay, G. and I. Bayramoglu. 2015. *Joint Distribution of New Sample Rank of Bivariate Order Statistics*. Journal of Applied Statistics, Vol. 42(10): 2280–2289.

[48] Kim, S.H. and H.A. David. (1990). *On The Dependence Structure of Order Statistics and Concomitants of Order Statistics*. J. Statist. Plann. Inference, 24: 363–368.

[49] Lai, C.D. 1978. *Morgenstern's Bivariate Distibution and Its Application to Point Process*. Journal of Mathematical Analysis and Applications, Vol. 65(2): 247–256.

[50] Lehmann, E.L. 1966. *Some Consepts of Dependence*. Ann. Math. Statist., Vol. 37(5): 1137–1153.

[51] Louzada, F. vd. 2012. *The Long-Term Bivariate Survival FGM Copula Model: An Application to a Brazilian HIV Data*. Journal of Data Science Vol. 10: 511-535.

[52] Matveychuk, S.A. and Yh.I. Petunin. 1990. *A Generalization of the Bernoulli Model Arising in Order Statistics*. I. Ukrainian Math. J., Vol. 42: 518–528.

[53] Matveychuk, S.A. and Yh.I. Petunin. 1991. *A Generalization of the Bernoulli Model Arising in Order Statistics*. II. Ukrainian Math. J., Vol. 43: 779–785.

[54] Nelsen, R.B. vd. 2003. *Kendall Distribution Functions*. Statistics & Probability Letters, Vol. 65(3): 263–268.

[55] Nelsen, R.B. 2006. *An Introduction to Copulas* (2^{nd} edition). USA: Springer.

[56] Nevzorov, V.B. 2001. *Records: Mathematical Theory.* American Mathematical Society, Providence, RI.

[57] Robbins, H. 1944. *On Distribution-Free Tolerance Limits in Random Sampling*. Ann. Math. Statistics, Vol. 15(2): 214–216.

[58] Sarkadi K. 1957. *On the Distribution of the Number of Exceedances*. Ann. Math Statist. Math, Vol. 28(4): 1021–1023.

[59] Stoimenova, E. and N. Balakrishnan. 2011. *A Class of Exceedance-Type Statistics for the Two-Sample Problem.* Journal of Statistical Planning and Inference, Vol. 141(9): 3244–3255.

[60] Stepanov, A. 2004. *Random Intervals Based on Records*. Journal of Statistical Planning and Inference, Vol. 118(1-2): 103–113.

[61] Su, J., N. Su, and W. Huang. 2008. *Characterizations Based on Record Values and Order Statistics*. Journal of Statistical Planning and Inference, Vol. 138(5): 1358 – 1367.

[62] Turhan, N.O. 2012. *On Asymptotic Distribution of Exceedance Statistics*. Journal of Computational and Applied Mathematics, Vol. 236(9): 2259–2267.

[63] Urden, L.D., K.M. Stacy and M.E. Lough. 2014. *Critical Care Nursing: Diagnosis and Management* (7^{th} edition). Canada: Elsevier.

[64] Vaughan, R.J. and W.N. Venables. 1972. *Permanent Expressions for Order Statistics Densities*. Journal of the Royal Statistical Society: Series B (Methodological), Vol. 34(2): 308–310.

[65] Yang, S.S. 1977. *General Distribution Theory of the Concomitants of Order Statistics*. The Annals of Statistics, Vol. 5(5): 996–1002.

[66] Wesolowski, J. and M. Ahsanullah. 1998. *Distributional Properties of Exceedance Statistics*. Ann. Inst. Statist. Math., Vol. 50(3): 543–565.

[67] Westcott, M. 1977a. *A Note on Record Times*. J. Appl. Probability, Vol. 14(3): 637–639.

[68] WHO (World Health Organization). 2006. *WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide: Global Update 2005. Summary of Risk Assestment*. Geneva: World Health Organization.

[69] Wilks, S.S. 1941. *Determination of sample sizes for setting tolerance limits*. Ann. Math. Statistics, Vol. 12(1): 91–96.

[70] Wilks, S.S. 1942. *Statistical Prediction with Special Reference to the Problem of Tolerance Limits*. Ann. Math. Statistics, Vol. 13(4): 400–409.

VITA

Ayşegül Erem was born in Izmir, Turkey, on November 3, 1989. She got her bachelor' s degree in Statistics from the Ege University in June 2012 with a degree of 3.59/4.00. After that she started her PhD education in Izmir University of Economics, Turkey.

Her research interests include applied probability and statistics, distribution theory, theory of statistics and copulas. She has given several talks in international conferences devoted to theory of statistics and applied statistics.

Since 2012, she has been awarded by The Scientific and Technological Research Council of Turkey (TÜBİTAK) with a PhD scholarship.