ORDER STATISTICS FROM NONIDENTICALLY DISTRIBUTED RANDOM VARIABLES AND EXCEEDANCES

GÖKNUR GINER

JUNE 2013

ORDER STATISTICS FROM NONIDENTICALLY DISTRIBUTED RANDOM VARIABLES AND EXCEEDANCES

A DISSERTATION SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL
AND APPLIED SCIENCES OF
IZMIR UNIVERSITY OF ECONOMICS

BY
GÖKNUR GINER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

JUNE 2013

Ph.D. DISSERTATION EXAMINATION RESULT FORM

We have read the dissertation entitled **"ORDER STATISTICS FROM NONIDENTICALLY DISTRIBUTED RANDOM VARIABLES AND EXCEEDANCES"** completed by **Göknur Giner** 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

Assoc. Prof. Dr. Selma Gürler
Thesis Committee Member

Asst. Prof. Dr. Güvenç Arslan
Thesis Committee Member

Prof. Dr. Ünal Ufuktepe
Thesis Committee Member

Assoc. Prof. Dr. Özlem Ege Oruç
Thesis Committee Member

Prof. Dr. Murat Aşkar
Director

ABSTRACT

ORDER STATISTICS FROM NONIDENTICALLY DISTRIBUTED RANDOM VARIABLES AND EXCEEDANCES

Göknur Giner
Ph.D. in Applied Mathematics and Statistics
Graduate School of Natural and Applied Sciences
Supervisor: Prof. Dr. İsmihan Bayramoğlu
June 2013

Order statistics and exceedances for some general models of independent but not necessarily identically distributed (INID) random variables are considered. The distributions of order statistics from INID sample are described in terms of symmetric functions. Some exceedance models based on order statistics from INID random variables are considered, the limit distributions of exceedance statistics are obtained. For the model of INID random variables referred as F^{α}-scheme introduced by [62] the limiting distribution of exceedance statistics has been derived. This distribution is expressed in terms of permutations with inversions, Gaussian hypergeometric function and incomplete beta function. Some applications in insurance models have also been discussed.

Keywords: Order statistics, INID random variables, exceedances, symmetric functions, Gaussian hypergeometric distribution, permutations with inversions.

ÖZ

FARKLI DAĞILIMLARA SAHİP RASGELE DEĞİŞKENLERİN SIRA İSTATİSTİKLERİ VE AŞAN İSTATİSTİKLER

Göknur Giner
Uygulamalı Matematik ve İstatistik, Doktora
Fen Bilimleri Enstitüsü
Tez Yöneticisi: Prof. Dr. İsmihan Bayramoğlu
Haziran 2013

Bağımsız ve farklı dağılımlara sahip rasgele değişkenlerin bazı modelleri için sıra istatistikleri ve aşan istatistikler kullanıldı. Ayrıca farklı dağılıma sahip rasgele değişkenlerin sıra istatistiklerine ait dağılımlar simetrik fonksiyonlar yardımıyla tanımlandı. Farklı dağılımlara sahip rasgele değişkenlerin sıra istatistikleri temel alınan aşan istatistiklerin asimptotik dağılımları elde edildi. Söz konusu dağılım bulunurken ilk defa [62] tarafından önerilmiş olan F^{α}-düzeni kullanıldı. Bu dağılım tersinmeli permütasyonlar, Gauss hipergeometrik ve tamamlanmamış beta fonksiyonları ile ifade edildi. Son olarak da sigortacılık modelleri ile igili bazı uygulama önerilerinden söz edildi.

Anahtar Kelimeler: Sıra istatistikleri, farklı dağılıma sahip rasgele değişkenler, aşan istatistikler, simetrik fonksiyonlar, Gauss hipergeometrik dağlımı, tersinmeli permütasyonlar.

ACKNOWLEDGEMENT

It is the end of a long, difficult and rewarding journey and I would like to thank everyone who has made it possible.

First and foremost, I would like to express my special thanks to my supervisor, Prof. Dr. İsmihan Bayramoğlu. I am grateful for his academic guidance, advice, patience and many interesting discussions on all topics, scientific or otherwise. He has inspired me with his wisdom and knowledge and helped me see ways to judge real life problems using the theory of statistics. It has been an honour and privilege to work with him.

I would like to acknowledge Assoc. Prof. Dr. Selma Gürler and Asst. Prof. Dr. Güvenç Arslan for their very helpful assistance and feedback in my committee. I also would like to thank Assoc. Prof. Dr. Gözde Yazgı Tütüncü, Sinan Kapçak and Gülder Kemalbay for their dearest friendships, academic advice which improved this study and priceless discussions on mathematics. I thank them all for many great memories, and my very special thanks to Sinan Kapçak for his selection of sophisticated music; Frédéric Chopin motivated me more than anything during my studies.

I am extremely grateful to Izmir University of Economics for providing me a scholarship during my PhD.

I also want to thank all the past and present colleagues at the department of Mathematics, especially Kerem Türkyılmaz, Hande Tunçel, Cihangir Kan, Femin Yalçın, Murat Özkut ve Ceki Franko who were there for me throughout my studies and who would listen to my rants and cheer me up when I was overwhelmed by it all. It was a great pleasure to work with them all.

I would like to thank my family. A huge thank you to my father Muammer Yapakçı for his wisdom, for being my inspiration and giving me insight into understanding the world in a scientific way since I was a little girl. Thank you to my mother Nurşen Yapakçı for remembering me in her prayers and special thanks to my friend, more than a friend, a sister for me, İdil Öztamur. They have always believed in my abilities and their support and encouragement have helped me in difficult times.

Last but by no means least, I am very grateful to my dear husband Rick Giner and his family for their continuous support and encouragement. I would like to thank him for his counsel and endless patience throughout this journey. His unconditional love has given me complete happiness.

To the loving memory of my dear brother Gökmert

TABLE OF CONTENTS

Chapter 1

Introduction

This dissertation centers around the problem of obtaining the distribution of independent and non-identically distributed (INID) random variables. The study presented deals with the main difficulties of the given problem and also suggests a useful method which makes the calculations easier. With the help of this method which is called F^{α} scheme first presented by [62], we obtained the asymptotic distribution function of exceedances. An overview of the other relevant paper and researches are useful to note that at this point.

The developments on order statistics compiled in [73] through early 1960's. The book [73] was primarily the contributions of several tables. Comprehensive tables have also been prepared by [40]. Based on their growing amount of importance and popularity, a considerable amount of research has investigated the order statistics. There have been books written on many different aspects of them. The first essential book describing the theory of order statistics is [28]. This book was the culmination of H. A. David's work from 1971. [2] and [29] include new developments on order statistics from independent and identically distributed (IID), INID and dependent random variables. [20] and [21] are two volumes on theory and applications of order statistics. Books written by [52], [34], [70], [69] and [30] first presented the asymptotic theory of order statistics. The theory and applications of order statistics from IID random variables are both studied well. However, there are not many developments on the theory of order statistics from arbitrarily dependent random variables because of difficulties in calculating the joint probability density function (pdf)'s. They are not factorized as they are in the case of IID. The marginal distribution function of an order statistic from arbitrary dependent random variables is given in [29]. The joint distribution function of two or more order statistics from dependent random variables can be found in [60]. The distribution theory of order statistics from INID random variables first mentioned in [78] and it involves the concept of permanents which is similar to the determinant except that the permanent does not have an alternating sign. The signs of all the terms in the summation given in the definition of the determinant are positive in the calculation of permanent. [17] is an excellent review which gives the theory of order statistics from INID case. [39] is also a study in the mean residual life functions for INID random variables in the system level.

Another key concept considered in this dissertation is exceedance statistics which denotes the total number of observations exceeding a random threshold value. The concept of a random threshold was first used by [38] and [31]. In this context, exceedance statistics have received a growing amount of attention among researchers like [71], [74], [82] and [75] and even earlier in [80], [81]. First fundamental source in the area is the chapter "The distribution of exceedance" in the book [36]. Further discussions can be found in [28], [48], [79] and [52]. Based on exceedance statistics, several theoretical studies are contributed to the area. Papers that examine the exceedance statistics in record model are [3], [5],[6], [13], [12] and the book is [4]. Exceedance models in multivariate FGM are introduced in the work of [7] and also [8] used exceedances in the progressive type II censoring scheme. The reader may refer to the literature [14] and [15] for the discussion on the distributions of exceedances of generalized order statistics. In [9] the joint behaviour of precedence and exceedances in random threshold models are obtained and lastly, [10] addresses the waiting times of exceedance statistics in random threshold models.

In Chapter 1, the fundamental definitions and distributional properties of the order statistics from IID and arbitrarily dependent random variables are given.

In Chapter 2, INID random variables are considered, and the distributions of order statistics are described in terms of permanents and the symmetric functions. Permanents, symmetric functions and their properties are also given, and the recurrence relations between the distributions of order statistics from INID random variables are provided.

In Chapter 3, we introduce the exceedance statistics which is one of the main concepts of this study. We extensively review the literature on exceedance statistics which includes the distribution of IID and INID random variables and arbitrarily dependent random variables with respect to a random threshold. Various exceedance models are included in the last section of the chapter.

In Chapter 4, the major contribution of the dissertation, the exceedance model from INID random variables is considered. The asymptotic distribution of given exceedance statistic has been derived, and for a special model of INID random variables the limiting distribution is studied. Behaviour of the numerical characteristics of derived limiting distribution, such as mean, variance and skewness are also interpreted. We focus on how the methodological considerations have been put into practice in insurance models. Lastly, chapter 5, the concluding remarks are given in the final chapter.

1.1 Order statistics from IID random variables

Suppose that $X_1, X_2, \dots, X_n$ are independent and identically distributed random variables with absolutely continuous cumulative distribution function $F(x)$ and $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ are the corresponding order statistics. The joint pdf of $X_{r_1:n}, X_{r_2:n}, \dots, X_{r_k:n}$ $(1 \le r_1 < r_2 \cdots < r_k \le n)$ is for $x_1 \le x_2 \le \cdots \le x_k$,

$$\begin{aligned} & f_{r_1,r_2,\dots,r_k}(x_1, x_2, \dots, x_k) \\ &= \frac{n!}{(r_1-1)!(r_2-r_1-1)!\cdots(n-r_k)!} \\ & F^{r_1-1}(x_1)f(x_1)[F(x_2)-F(x_1)]^{r_2-r_1-1}f(x_2) \\ & [F(x_3)-F(x_2)]^{r_3-r_2-1}f(x_3)\cdots[1-F(x_k)]^{n-r_k}f(x_k) \end{aligned} \tag{1.1}$$

Defining $x_0 = -\infty, x_{k+1} = +\infty, n_0 = 0, r_0 = 0, r_{k+1} = n+1$, we can simplify the equation (1.1) as

$$\begin{aligned} & f_{r_1,r_2,\dots,r_k}(x_1, x_2, \dots, x_k) \\ &= \left[n! \prod_{j=1}^{k} f(x_j)\right] \prod_{j=0}^{k} \left\{ \frac{[F(x_{j+1})-F(x_j)]^{r_{j+1}-r_j-1}}{(r_{j+1}-r_j-1)!} \right\} \end{aligned} \tag{1.2}$$

Since there are $n!$ equally likely orderings of the x_i, it is clear that the joint pdf of all n order statistics is

$$\begin{aligned} f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) &= n! f(x_1) f(x_2) \dots f(x_n) \\ &= n! \prod_{i=1}^{n} f(x_i), \quad x_1 \le x_2 \le \cdots \le x_n \end{aligned} \tag{1.3}$$

Let $f_{r,s}(x, y)$ denotes the joint pdf of $X_{r:n}$ and $X_{s:n}$ $(1 \le r < s \le n)$.

Then for $x \leq y$, we can obtain $f_{r,s:n}(x,y)$ by integrating out (1.3) the variables $(X_{1:n},\dots,X_{r-1:n})$, $(X_{r+1:n},\dots,X_{s-1:n})$ and $(X_{s+1:n},\dots,X_{n:n})$

$$\begin{aligned} f_{r,s:n}(x,y) &= n!f(x)f(y)\left\{\int_{-\infty}^{x}\cdots\int_{-\infty}^{x_2} f(x_1)\cdots f(x_{r-1})dx_1\cdots dx_{r-1}\right\} \\ &\quad\times\left\{\int_{x}^{y}\cdots\int_{x}^{x_{r+2}} f(x_{r+1})\cdots f(x_{s-1})dx_{r+1}\cdots dx_{s-1}\right\} \\ &\quad\times\left\{\int_{y}^{\infty}\cdots\int_{y}^{x_{s+2}} f(x_{s+1})\cdots f(x_n)dx_{s+1}\cdots dx_n\right\} \\ &= n!f(x)f(y)\frac{[F(x)]^{r-1}}{(r-1)!}\times\frac{[F(y)-F(x)]^{s-r-1}}{(s-r-1)!}\times\frac{[1-F(y)]^{n-s}}{(n-s)!} \\ &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}[F(x)]^{r-1}[F(y)-F(x)]^{s-r-1} \\ &\quad [1-F(y)]^{n-s}f(x)f(y), \qquad (1.4) \end{aligned}$$

$$-\infty < x_1 < \cdots < x_{r-1} < x < x_{r+1} < \cdots < x_{s-1} < y < x_{s+1} < \cdots < x_n < \infty.$$

By letting $r = 1$ and $s = n$, the joint distribution function of the first and the largest order statistics becomes

$$f_{1,n}(x,y) = n(n-1)[F(y)-F(x)]^{n-2}f(x)f(y), \quad -\infty < x < y < \infty. \qquad (1.5)$$

For $x < y$, the joint cdf of $X_{r:n}$ and $X_{s:n}$, denoted by $F_{r,s}(x,y)$ can be obtain by integration of $f_{r,s}(x,y)$ as follows:

$$\begin{aligned} F_{r,s}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y}\frac{n!}{(r-1)!(s-r-1)!(n-s)!}F^{r-1}(x)[F(y)-F(x)]^{s-r-1} \\ &\quad [1-F(y)]^{n-s}dF(x)dF(y) \qquad (1.6) \end{aligned}$$

For discrete case, we have for $x < y$

$$\begin{aligned} F_{r,s}(x,y) &= P\{\text{at least}\, r \,\text{of}\, X_i\text{'s} \leq x, \text{at least}\, s \,\text{of}\, X_i\text{'s} \leq y\} \qquad (1.7) \\ &= \sum_{j=s}^{n}\sum_{i=r}^{j} P\{\text{exactly}\, i \,\text{of}\, X_i\text{'s} \leq x, \text{at least}\, j \,\text{of}\, X_i\text{'s}\ \leq y\} \\ &= \sum_{j=s}^{n}\sum_{i=r}^{j} \frac{n!}{i!(j-i)!(n-j)!} F^i(x)[F(y)-F(x)]^{j-i}[1-F(y)]^{n-j}, \end{aligned}$$

and for $x \geq y$

$$F_{r,s}(x,y) = F_s(y).$$

We can obtain the marginal distribution of $X_{r:n}$ $(1 \leq r \leq n)$, say $f_{r:n}(x)$, by integrating out (1.3) the variables $(X_{1:n}, \ldots, X_{r-1:n})$ and $(X_{r+1:n}, \ldots, X_{n:n})$

$$\begin{aligned} f_{r:n}(x) &= n! f(x) \left\{ \int_{-\infty}^{x} \cdots \int_{-\infty}^{x_2} f(x_1)\cdots f(x_{r-1}) dx_1 \cdots dx_{r-1} \right\} \\ &\quad \times \left\{ \int_{x}^{\infty} \cdots \int_{x}^{x_{r+2}} f(x_{r+1})\cdots f(x_n) dx_{r+1}\cdots dx_n \right\} \\ &= n! f(x) \frac{[F(x)]^{r-1}}{(r-1)!} \times \frac{[1-F(x)]^{n-r}}{(n-r)!} \\ &= \frac{n!}{(n-r)!(r-1)!}[F(x)]^{r-1}[1-F(x)]^{n-r} f(x), \qquad (1.8) \end{aligned}$$

$-\infty < x_1 < \cdots < x_{r-1} < x < x_{r+1} < \cdots < x_n < \infty.$

Distribution of the range statistics

Range is the difference between the largest and the smallest observation in the sample, i.e. the special case of spacing $Z_{rs} = X_{s:n} - X_{r:n}$ $(1 \leq r < s \leq n)$ where $r = 1$ and $s = n$.

Let us start with the pdf of Z_{rs}. To find the pdf of $Z_{rs} = X_{s:n} - X_{r:n}$ we put $z_{rs} = y - x$ in (1.4) and the Jacobian is unit in modulus for transformation from x, y to x, z_{rs}. Hence we get

$$f_{r,s}(x, z_{rs}) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \int_{-\infty}^{\infty} F^{r-1}(x)[F(x+z_{rs}) - F(x)]^{s-r-1}$$

$$f(x+z_{rs})[1 - F(x+z_{rs})]^{n-s} dF(x).$$

Therefore, for the case $r = 1, s = n$, Z_{rs} becomes the range statistics. Let us denote $Z_{1n} = R$, thus the pdf of R can be obtained as follows:

$$f_R(t) = n(n-1) \int_{-\infty}^{\infty} [F(x+t) - F(x)]^{n-2} f(x+t) dF(x). \tag{1.9}$$

The cdf of R is

$$\begin{aligned} F_R(t) &= n \int_{-\infty}^{\infty} f(x) \int_{0}^{t} (n-1) f(x+t^{'})[F(x+t^{'}) - F(x)]^{n-2} dt^{'} dx \\ &= n \int_{-\infty}^{\infty} f(x)[F(x+t^{'}) - F(x)]^{n-1} \Big|_{t^{'}=0}^{t^{'}=t} dx \qquad (1.10) \\ &= n \int_{-\infty}^{\infty} [F(x+t) - F(x)]^{n-1} dF(x) \end{aligned}$$

1.2 Order Statistics from dependent random variables

Let $X_1, X_2, \ldots, X_n$ be arbitrarily dependent continuous random variables with joint pdf $f(x_1, x_2, \ldots, x_n)$, $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. The joint pdf of the corresponding order statistics $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ can be easily obtained by considering the limit of

$$\frac{1}{\delta x_1 \cdots \delta x_n} P\{x_1 < X_{1:n} < x_1 + \delta x_1, \ldots, x_n < X_{n:n} < x_n + \delta x_n\}$$

$$= \frac{1}{\delta x_1 \cdots \delta x_n} \sum_{i_1, i_2, \ldots, i_n} P\{x_1 < X_{i_1} < x_1 + \delta x_1, \ldots,$$

$$x_n < X_{i_n} < x_n + \delta x_n, X_{i_1} < X_{i_2} < \cdots < X_{i_n}\}$$

as $\delta x_i \rightarrow 0, i = 1, 2, .., n$. Let $f_{X_{1:n}, X_{2:n}, \ldots, X_{n:n}}$ denotes the joint pdf of $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$, then

$$f_{X_{1:n}, X_{2:n}, \ldots, X_{n:n}}(x_1, x_2, \ldots, x_n)$$

$$= \begin{cases} \sum\limits_{\wp_{1,2,\ldots,n}} f(x_{i_1}, x_{i_2}, \ldots, x_{i_n}) & \text{if} \quad x_1 < x_2 < \cdots < x_n \\ 0 & \text{otherwise} \end{cases}, \tag{1.11}$$

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

If $X_1, X_2, \ldots, X_n$ are exchangeable, i.e. $f(x_1, x_2, \ldots, x_n) = f(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$, $\forall (i_1, i_2, \ldots, i_n) \in \wp_{1,2,\ldots,n}$ and $f_{X_j}(x) = f_X(x)$, $j = 1, 2 \ldots, n$, then we have

$$f_{X_{1:n}, X_{2:n}, \ldots, X_{n:n}}(x_1, x_2, \ldots, x_n)$$

$$= \begin{cases} n! f(x_1, x_2, \ldots, x_n) & \text{if} \quad x_1 < x_2 < \cdots < x_n \\ 0 & \text{otherwise} \end{cases} \tag{1.12}$$

The marginal pdf's of order statistics $X_{k_1:n}, X_{k_2:n}, \ldots, X_{k_r:n}$, $1 \leq k_1 < k_2 < \cdots < k_r \leq n$ in case of arbitrarily dependent random variables can be obtained by integrating (1.11) as follows:

$$f_{X_{k_1:n},X_{k_2:n},\ldots,X_{k_r:n}}(x_{k_1},x_{k_2},\ldots,x_{k_r})$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{x_{n-1}}\cdots\int_{-\infty}^{x_{k_r}}\int_{-\infty}^{x_{r-1}}\cdots\int_{-\infty}^{x_{k_2+1}}\int_{-\infty}^{x_{k_2-1}}\cdots\int_{-\infty}^{x_{k_1+1}}\int_{-\infty}^{x_{k_1-1}}\cdots\int_{-\infty}^{x_3}\int_{-\infty}^{x_2}$$

$$f_{X_{1:n},X_{2:n},\ldots,X_{n:n}}(x_1,x_2,\ldots,x_n)\times dx_1dx_2\cdots dx_{k_1-1}$$

$$\times dx_{k_1+1}\cdots dx_{k_r-1}dx_{k_r+1}\cdots dx_n$$

$$= \sum_{\wp_{1,2,\ldots,n}}\int_{-\infty}^{\infty}\int_{-\infty}^{x_{n-1}}\cdots\int_{-\infty}^{x_{k_r}}\int_{-\infty}^{x_{r-1}}\cdots\int_{-\infty}^{x_{k_2+1}}\int_{-\infty}^{x_{k_2-1}}\cdots\int_{-\infty}^{x_{k_1+1}}\int_{-\infty}^{x_{k_1-1}}\cdots\int_{-\infty}^{x_3}\int_{-\infty}^{x_2} \quad (1.13)$$

$$f(x_{i_1},x_{i_2},\ldots,x_{i_n})\times dx_1dx_2\cdots dx_{k_1-1}dx_{k_1+1}\cdots dx_{k_r-1}dx_{k_r+1}\cdots dx_n.$$

The marginal pdf's of order statistics $X_{k_1:n},X_{k_2:n},\ldots,X_{k_r:n}$, $1\le k_1<k_2<\cdots<k_r\le n$ in case of exchangeable random variables can be obtained by integrating (1.12) as follows:

$$f_{X_{k_1:n},X_{k_2:n},\ldots,X_{k_r:n}}(x_{k_1},x_{k_2},\ldots,x_{k_r})$$

$$= n!\int_{-\infty}^{\infty}\int_{-\infty}^{x_{n-1}}\cdots\int_{-\infty}^{x_{k_r}}\int_{-\infty}^{x_{r-1}}\cdots\int_{-\infty}^{x_{k_2+1}}\int_{-\infty}^{x_{k_2-1}}\cdots\int_{-\infty}^{x_{k_1+1}}\int_{-\infty}^{x_{k_1-1}}\cdots\int_{-\infty}^{x_3}\int_{-\infty}^{x_2} \quad (1.14)$$

$$f(x_1,x_2,\ldots,x_n)\times dx_1dx_2\cdots dx_{k_1-1}dx_{k_1+1}\cdots dx_{k_r-1}dx_{k_r+1}\cdots dx_n.$$

Chapter 2

Order statistics from INID random variables

Definitions and properties of order statistics from INID random variables and permanents presented in this chapter mainly refer to work in [17], [29] and [2]. In addition, we give the definitions and the recurrence relations of the order statistics from INID random variables in the meaning of symmetric functions.

Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables with the cumulative distribution functions (cdf) $F_1(x), F_2(x), \ldots, F_n(x)$ and the probability density functions $f_1(x), f_2(x), \ldots, f_n(x)$, $\forall x \in \mathbb{R}$ respectively. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the corresponding order statistics.

We can show that the joint density function of $X_{r_1:n}, X_{r_2:n}, \ldots, X_{r_k:n}$ which is denoted by $f_{r_1,r_2,\ldots,r_k}(x_1, x_2, \ldots, x_k)$ for $1 \le r_1 < r_2 < \cdots < r_k \le n$ is

$$\begin{aligned} & f_{r_1,r_2,\ldots,r_k}(x_1, x_2, \ldots, x_k) \\ &= \frac{1}{(r_1-1)!(r_2-r_1-1)!\cdots(r_k-r_{k-1}-1)!(n-r_k)!} \sum_{\wp_{1,2,\ldots,n}} \prod_{j=1}^{k} f_{i_{r_j}}(x_j) \prod_{j=1}^{r_1-1} F_{i_j}(x_1) \\ & \quad \times \prod_{j=r_1+1}^{r_2-1} [F_{i_j}(x_2) - F_{i_j}(x_1)] \cdots \prod_{j=r_{k-2}+1}^{r_{k-1}-1} [F_{i_j}(x_{k-1}) - F_{i_j}(x_{k-2})] \\ & \quad \times \prod_{j=r_{k-1}+1}^{n-r_k-1} [1 - F_{i_j}(x_k)], \qquad (2.1) \end{aligned}$$

$-\infty < x_1 < x_2 < \cdots x_k < \infty$ and where $\sum_{\wp_{1,2,\ldots,n}}$ denotes the sum over all $n!$ permutations $(j_1, j_2, \ldots, j_n)$ of $(1, 2, \ldots, n)$.

Let $F_{r_1,r_2,\ldots,r_k}(x_1, x_2, \ldots, x_k)$ denotes the cdf of $X_{r_1:n}, X_{r_2:n}, \ldots, X_{r_k:n}$ for $1 \le r_1 < r_2 < \cdots < r_k \le n$. Then $F_{r_1,r_2,\ldots,r_k}(x_1, x_2, \ldots, x_k)$ can be expressed as follows

$$\begin{aligned} & F_{r_1,r_2,\ldots,r_k}(x_1, x_2, \ldots, x_k) \\ &= \sum \frac{1}{j_1! j_2! \cdots j_{k+1}!} \prod_{j=1}^{k} f_{i_{r_j}}(x_j) \prod_{j=1}^{r_1-1} F_{i_j}(x_1) \\ & \quad \times \prod_{j=r_1+1}^{r_2-1} [F_{i_j}(x_2) - F_{i_j}(x_1)] \cdots \prod_{j=r_{k-2}+1}^{r_{k-1}-1} [F_{i_j}(x_{k-1}) - F_{i_j}(x_{k-2})] \\ & \quad \times \prod_{j=r_{k-1}+1}^{n-r_k-1} [1 - F_{i_j}(x_k)], \qquad (2.2) \end{aligned}$$

$-\infty < x_1 < x_2 < \cdots < x_k < \infty$ and where the sum is over all $j_1, j_2, \ldots, j_{k+1}$ with $j_1 \ge r_1, j_1 + j_2 \ge r_2, \ldots, \sum_{i=1}^{k} j_i \ge r_k$ and $j_1 + j_2 + \cdots + j_k = n$.

The joint probability density function of all n order statistics can be obtained from (2.1) as

$$f_{X_{1:n},X_{2:n},\ldots,X_{n:n}}(x_1,x_2,\ldots,x_n)$$

$$= \begin{cases} \sum\limits_{\wp_{1,2,\ldots,n}} f_{i_1}(x_1)f_{i_2}(x_2)\cdots f_{i_n}(x_n) & -\infty < x_1 < x_2 < \cdots < x_n < \infty \\ 0 & \text{otherwise} \end{cases}$$

or it can be written explicitly in the form

$$f_{X_{1:n},X_{2:n},\ldots,X_{n:n}}(x_1,x_2,\ldots,x_n)$$

$$= \begin{cases} \sum\limits_{\wp_{1,2,\ldots,n}} \prod\limits_{j=1}^{n} f_{i_j}(x_j) & -\infty < x_1 < x_2 < \cdots < x_n < \infty \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

where $\sum\limits_{\wp_{1,2,\ldots,n}}$ denotes the sum over all $n!$ permutations $(j_1, j_2, \ldots, j_n)$ of $(1, 2, \ldots, n)$.

Similarly, the following joint density function for $X_{r:n}$ and $X_{s:n}$ $(1 \le r < s \le n)$ may be defined as

$$\begin{aligned} f_{r,s:n}(x,y) &= \frac{1}{(r-1)!(s-r-1)!(n-s)!} \sum_{\wp_{1,2,\ldots,n}} F_{i_1}(x)\cdots F_{i_{r-1}}(x) f_{i_r}(x) \\ &\quad \times [F_{i_{r+1}}(y) - F_{i_{r+1}}(x)]\cdots[F_{i_{s-1}}(y) - F_{i_{s-1}}(x)] \\ &\quad \times f_{i_s}(x) \times [1 - F_{i_{s+1}}(y)]\cdots[1 - F_{i_n}(y)], \end{aligned}$$

$-\infty < x < y < \infty$ or explicitly,

$$\begin{aligned} f_{r,s:n}(x,y) &= \frac{1}{(r-1)!(s-r-1)!(n-s)!} \sum_{\wp_{1,2,\ldots,n}} f_{i_r}(x) f_{i_s}(x) \prod_{j=1}^{r-1} F_{i_j}(x) \\ &\quad \times \prod_{j=r+1}^{s-1} [F_{i_j}(y) - F_{i_j}(x)] \prod_{j=s+1}^{n} [1 - F_{i_j}(y)], \end{aligned} \tag{2.4}$$

$-\infty < x < y < \infty$.

The joint cumulative distribution function of $X_{r:n}$ and $X_{s:n}$ is

$$\begin{aligned} F_{r,s}(x,y) &= \sum_{i=r}^{n} \sum_{j=\max(0,s-i)}^{n-i} \frac{1}{(i-1)!(j-i-1)!(n-j)!} \sum_{\wp_{1,2,\ldots,n}} \prod_{l=1}^{i} F_{t_l}(x) \times \\ &\quad \times \prod_{l=i+1}^{i+j} (F_{t_l}(y) - F_{t_l}(x)) \prod_{l=i+j+1}^{n} (1 - F_{t_l}(y)), \qquad (2.5) \end{aligned}$$

where the summation $\wp_{1,2,\ldots,n}$ extends over all $n!$ permutations $(t_1, t_2, \ldots, t_n)$ of $(1, 2, \ldots, n)$.

We can derive the pdf and the cdf of $X_{r:n}$ for $1 \leq r \leq n$ as follows respectively,

$$\begin{aligned} f_{r:n}(x) &= \frac{1}{(r-1)!(n-r)!} \sum_{\wp_{1,2,\ldots,n}} F_{i_1}(x) \cdots F_{i_{r-1}}(x) f_{i_r}(x) \\ &\quad \times [1 - F_{i_{r+1}}(x)] \cdots [1 - F_{i_n}(x)] \qquad (2.6) \end{aligned}$$

and

$$\begin{aligned} F_{r:n}(x) &= P\{X_{r:n} \leq x\} \\ &= P\{\text{at least}\, r \,\text{of}\, X'\text{s are} \leq x\} \\ &= \sum_{i=r}^{n} P\{\text{exactly}\, i \,\text{of}\, X'\text{s are} \leq x\} \\ &= \sum_{i=r}^{n} \frac{1}{i!(n-i)!} \sum_{\wp_{1,2,\ldots,n}} \prod_{l=1}^{i} F_{j_l}(x) \prod_{l=i+1}^{n} [1 - F_{j_l}(x)], \qquad (2.7) \end{aligned}$$

where $\sum_{\wp_{1,2,\ldots,n}}$ denotes the sum over all $n!$ permutations $(j_1, j_2, \ldots, j_n)$ of $(1, 2, \ldots, n)$.

2.1 Order statistics based on permanents

The distribution theory of order statistics from INID random variables involving permanents made its first appearance in [78]. [17] also gives a comprehensive review of the recent developments on this subject based on permanents.

2.1.1 Permanents

Permanent function is investigated by [26] and [24] in their respected memoirs. Since then, many mathematicians contributed to the subject. The monograph on permanents written by [42] is a superior reference for anyone interested in theory of permanents. In [26], it is distinguished determinants as alternating symmetric functions from determinants as ordinary symmetric functions. In the same paper, it is also introduced a subclass of symmetric functions given the name permanents by [58]. Permanent is a concept defined similar to the determinant except that it does not have an alternating sign, that is, no sign changes occur.

Definition 2.1. The permanent of an $n \times n$ square matrix $Z = (z_{km})$ is defined as

$$Per\mathbf{Z} = \left| \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ z_{n1} & z_{n2} & \cdots & z_{nn} \end{bmatrix} \right| = \sum_{\wp_{1,2,\ldots,n}} \prod_{m=1}^{n} z_{mk_m}, \tag{2.8}$$

where $\wp_{1,2,\ldots,n}$ is the class of all $n!$ permutations $(k_1, k_2, \ldots, k_n)$ of $(1, 2, \ldots, n)$ and the sum extends over all permutations.

2.1.2 Elementary properties of permanents

Permanents and determinants share many properties because of the similarity of their definition. However, permanents do not hold two key properties of determinants which are the multiplicative property and the invariance under certain elementary matrix operations. Permanent functions have the following fundamental properties.

Property 1 If the rows or columns of matrix $\mathbf{Z}$ are permuted, the value of $Per\mathbf{Z}$ does not change.

Property 2 The permanent of a matrix can be expanded by any row or column. Let $\mathbf{Z}_{k,m}$ denotes the sub-matrix of order $n-1$ obtained from $\mathbf{Z}$ by deleting the $k-th$ row and the $m-th$ column, then

$$\begin{aligned} Per\mathbf{Z} &= \sum_{k=1}^{n} z_{km} Per\mathbf{Z}_{k,m}, \qquad m = 1, 2, \ldots, n \\ &= \sum_{m=1}^{n} z_{km} Per\mathbf{Z}_{k,m}, \qquad k = 1, 2, \ldots, n. \end{aligned}$$

Property 3 Let c be a constant and $\mathbf{Z}_1$ denotes the matrix obtained from $\mathbf{Z}$ multiplying the elements in the $k-th$ row by c, $m = 1, 2, \ldots, n$, then

$$Per\mathbf{Z}_1 = cPer\mathbf{Z}.$$

Property 4 Let $\mathbf{Z}_2$ denotes the matrix obtained from $\mathbf{Z}$ by adding $t_{k,m}$ to the elements in the $k-th$ row, $m = 1, 2, \ldots, n$ and $\mathbf{Z}_1$ be the matrix obtained from $\mathbf{Z}$ by replacing the elements in the $k-th$ row by $t_{k,m}$, $m = 1, 2, \ldots, n$, then

$$Per\mathbf{Z}_2 = Per\mathbf{Z} + Per\mathbf{Z}_1.$$

2.1.3 Distributions of order statistics by permanents

Let $X_1, X_2, \ldots, X_n$ be independent random variables having the cumulative distributions $F_1(x), F_2(x), \ldots, F_n(x)$ and the probability density functions $f_1(x), f_2(x), \ldots, f_n(x)$, $\forall x \in \mathbb{R}$ respectively, and $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the corresponding order statistics.

From the identity (2.6) and (2.8) we can easily see that $f_{r:n}(x)$, the pdf of rth

orders statistics can be written in terms of permanents as

$$f_{r:n}(x) = \frac{1}{(r-1)!(n-r)!} Per Z_1, \forall x \in \mathbb{R} \tag{2.9}$$

where

$$Z_1 = \begin{bmatrix} F_1(x) & F_2(x) & \cdots & F_n(x) \\ f_1(x) & f_2(x) & \cdots & f_n(x) \\ \overline{F}_1(x) & \overline{F}_2(x) & \cdots & \overline{F}_n(x) \end{bmatrix} \begin{matrix} \}r-1 \\ \}1 \\ \}n-r \end{matrix}$$

and $\overline{F}_i(x) = 1 - F_i(x)$ for $i = 1, 2, \ldots, n$.

It is easy to express the cdf of rth order statistics given in (2.7) using permanents, as given in the following equation

$$F_{r:n}(x) = \sum_{i=r}^{n} \frac{1}{i!(n-i)!} Per T_1, \forall x \in \mathbb{R} \tag{2.10}$$

where

$$T_1 = \begin{bmatrix} F_1(x) & F_2(x) & \cdots & F_n(x) \\ \overline{F}_1(x) & \overline{F}_2(x) & \cdots & \overline{F}_n(x) \end{bmatrix} \begin{matrix} \}i \\ \}n-i \end{matrix}$$

On taking (2.4) and (2.8) into account we deduce the joint density $f_{r,s:n}(x, y)$ for $1 \le r < s \le n$ in the form as follows

$$f_{r,s:n}(x, y) = \frac{1}{(r-1)!(s-r-1)!(n-s)!} Per Z_2,\ \forall x, y \in \mathbb{R}, \tag{2.11}$$

where

$$Z_2 = \begin{bmatrix} F_1(x) & F_2(x) & \cdots & F_n(x) \\ f_1(x) & f_2(x) & \cdots & f_n(x) \\ F_1(y) - F_1(x) & F_2(y) - F_2(x) & \cdots & F_n(y) - F_n(x) \\ f_1(y) & f_2(y) & \cdots & f_n(y) \\ \overline{F}_1(y) & \overline{F}_2(y) & \cdots & \overline{F}_n(y) \end{bmatrix} \begin{matrix} \}r-1 \\ \}1 \\ \}s-r-1 \\ \}1 \\ \}n-s \end{matrix}$$

Similarly, from (2.1) the joint pdf of $X_{r_1:n}, X_{r_2:n}, \ldots, X_{r_k:n}$ for $1 \le r_1 < r_2 <$

$\cdots < r_k \leq n$ can be written based on permanents as

$$f_{r_1,r_2,\ldots,r_k}(x_1,x_2,\ldots,x_k) = \frac{1}{(r_1-1)!\cdots(r_k-r_{k-1}-1)!(n-r_k)!} Per Z_k, \quad (2.12)$$

$-\infty < x_1 < x_2 < \cdots, x_k < \infty$ and where

$$Z_k = \begin{bmatrix} F_1(x_1) & \cdots & F_n(x_1) \\ f_1(x_1) & \cdots & f_n(x_1) \\ F_1(x_2)-F_1(x_1) & \cdots & F_n(x_2)-F_n(x_1) \\ f_1(x_2) & \cdots & f_n(x_2) \\ \cdots & \cdots & \cdots \\ F_1(x_k)-F_1(x_{k-1}) & \cdots & F_n(x_k)-F_n(x_{k-1}) \\ f_1(x_k) & \cdots & f_n(x_k) \\ \overline{F}_1(x_k) & \cdots & \overline{F}_n(x_k) \end{bmatrix} \begin{matrix} \}r_1-1 \\ \}1 \\ \}r_2-r_1-1 \\ \}1 \\ \vdots \\ \}r_k-r_{k-1}-1 \\ \}1 \\ \}n-r_k \end{matrix}$$

We can show that the joint cdf of $X_{r_1:n}, X_{r_2:n}, \ldots, X_{r_k:n}$ for $1 \leq r_1 < r_2 < \cdots < r_k \leq n$

$$F_{r_1,r_2,\ldots,r_k}(x_1,x_2,\ldots,x_k) = \sum \frac{1}{j_1!j_2!\cdots j_{k+1}!} Per T_k, \quad (2.13)$$

$-\infty < x_1 < x_2 < \cdots, x_k < \infty$ and where

$$T_k = \begin{bmatrix} F_1(x_1) & \cdots & F_n(x_1) \\ F_1(x_2)-F_1(x_1) & \cdots & F_n(x_2)-F_n(x_1) \\ \cdots & \cdots & \cdots \\ F_1(x_k)-F_1(x_{k-1}) & \cdots & F_n(x_k)-F_n(x_{k-1}) \\ \overline{F}_1(x_k) & \cdots & \overline{F}_n(x_k) \end{bmatrix} \begin{matrix} \}j_1 \\ \}j_2 \\ \vdots \\ \}j_k \\ \}j_{k+1} \end{matrix}$$

and the sum is over $j_1, j_2, \ldots, j_{k+1}$ with $j_1 \geq r_1, j_1 + j_2 \geq r_2, \ldots \sum_{i=1}^{k} j_i \geq r_k$ and $j_1 + j_2 + \cdots + j_k = n$.

2.2 Order statistics based on symmetric functions

In this section, we present the distributions of order statistics from INID random variables based on the symmetric functions. Permanent expressions for the distribution function of INID order statistics allow us to gain some recurrence relations using the expansion of the permanent by some of the rows. However, in some cases when the applications of order statistics from the INID random variables are considered, the usage of the permanent expressions for the distributions of INID order statistics causes some difficulties connected with the complexity of operations. For instance, the mean residual life function of parallel and $k-out-of-n$ coherent systems when the life length of the components is INID random variables can not easily be calculated using permanent expressions. Therefore, the calculations involving the joint distributions of order statistics from INID random variables cover technical difficulties, the results are complicated and are not convenient for applications. The representations of distributions of order statistics from INID random variables in terms of symmetric functions have an advantage if one uses the derivatives and integration in calculations.

For this reason, we give the definition and the some properties of symmetric functions in the following subsection.

2.2.1 Symmetric functions

The books [54] and [76] can be used as guides for giving the following definitions.

Definition 2.2. Let $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ $(n \geq 1)$ and $S_r(\lambda_1, \ldots, \lambda_n)$ denotes the rth elementary symmetric function, where $1 \leq r \leq n$. $S_r(\lambda_1, \ldots, \lambda_n \geq n)$ is defined as

$$S_r(\lambda_1, \ldots, \lambda_n) = \sum_{1 \leq i_1 \leq \cdots \leq i_r \leq \cdots \leq i_n} \lambda_{i_1} \ldots \lambda_{i_n}, \tag{2.14}$$

that is, $S_r(\lambda_1, \ldots, \lambda_n)$ is the sum of all products of r distinct variables chosen from n variables.

Definition 2.3. Let $\Psi_r(\lambda_1, \ldots, \lambda_n)$ denotes the rth complete symmetric function, in variables $\lambda_1, \ldots, \lambda_n$.$\Psi_r(\lambda_1, \ldots, \lambda_n)$ is expressed as

$$\Psi_r(\lambda_1, \ldots, \lambda_n) = \sum_{\alpha_1+\ldots+\alpha_n=r} \lambda_1^{\alpha_1} \ldots \lambda_n^{\alpha_n}, \tag{2.15}$$

where the summation extends over all choices of $\alpha_1, \ldots, \alpha_n \in \{0, 1, \ldots, r\}$.

Both elementary and the complete symmetric functions satisfy the following condition

$$S_r(\lambda_1, \ldots, \lambda_n) = \Psi_r(\lambda_1, \ldots, \lambda_n) = \begin{cases} 1, & r < 0 \quad or \quad r > n \\ 0, & r = 0 \end{cases}$$

For other types of symmetric functions, one can see [76].

Generating functions for elementary and the complete symmetric functions can be defined as follows, respectively

$$\Phi(\lambda) = \prod_{i=1}^{n}(1 - \lambda_i\lambda) = \sum_{r=0}^{n}(-1)^r S_r(\lambda_1, \ldots, \lambda_n)\lambda^r \tag{2.16}$$

and

$$\frac{1}{\Phi(\lambda)} = \sum_{r=0}^{\infty}\Psi_r(\lambda_1, \ldots, \lambda_n)\lambda^r. \tag{2.17}$$

Since $\Phi(x)\dfrac{1}{\Phi(x)} = 1$, then

$$\sum_{r=0}^{n}(-1)^r S_r(\lambda_1, \ldots, \lambda_n)\Psi_{n-r}(\lambda_1, \ldots, \lambda_n) = 0,$$

$n \geq 1$.

The following recurrence relations of the symmetric functions may be obtained from (2.16) and (2.17) respectively

$$S_r(\lambda_1, \ldots, \lambda_n) = S_r(\lambda_1, \ldots, \lambda_{n-1}) + \lambda_n S_{r-1}(\lambda_1, \ldots, \lambda_{n-1}) \tag{2.18}$$

and

$$\Psi_r(\lambda_1, \ldots, \lambda_n) = \Psi_r(\lambda_1, \ldots, \lambda_{n-1}) + \lambda_n \Psi_{r-1}(\lambda_1, \ldots, \lambda_n) \qquad (2.19)$$

for $r \geq 1$ and $n \geq 2$.

2.2.2 Distributions of order statistics by symmetric functions

Let $X_1, X_2, \ldots, X_n$ be independent but not necessarily identically distributed random variables with cumulative distribution functions $F_1(x), F_2(x), \ldots, F_n(x)$ and $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ be corresponding order statistics.

For any borel set $B \in \Re$, where $\Re$ is the Borel $\sigma-$algebra of subsets of the set of real numbers $\mathbb{R}$, consider the indicator

$$I_{X_i}(B) = \begin{cases} 1, & X_i \in B \\ 0, & X_i \notin B \end{cases},$$

$i = 1, 2, \ldots, n$ and let $\nu^*(B) = \sum\limits_{i=1}^{n} I_{X_i}(B)$.

Define the empirical distribution of the INID sample $X_1, X_2, \ldots, X_n$ as $P_n^*(B) = \frac{\nu^*(B)}{n}$.

It is clear that

$$EI_{X_i}(B) = P\{X_i \in B\} = \int\limits_B dF_i(x) \equiv P_i(B)$$

and

$$Var(I_{X_i}(B)) = P_i(B)(1 - P_i(B)).$$

It's also obvious that $EP_n^*(B) = \frac{1}{n} \sum\limits_{i=1}^{n} P_i(B)$ and $Var(P_n^*(B)) = \frac{1}{n^2} \sum\limits_{i=1}^{n} P_i(B)(1 - P_i(B))$.

The empirical distribution function of the INID sample then is defined as

$$F_n^*(x) = P_n^*((-\infty, x]) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i}(x),$$

where $I_{X_i}(x) = 1$ if $X_i \leq x$ and $I_{X_i}(x) = 0$, otherwise.

According to the Kolmogorov's theorem the sequence of mutually independent random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ obeys the strong law of large numbers, if $\sum\limits_{n=1}^{\infty} \frac{Var(\xi_n)}{n^2} < \infty$ (see [35], Page 215).

Since

$$\frac{Var(I_{X_n}(B))}{n^2} = \frac{P_n(B)(1 - P_n(B))}{n^2} \leq \frac{1}{n^2}$$

then the series

$$\sum_{n=1}^{\infty} \frac{Var(I_{X_n}(B))}{n^2}$$

converges.

Then the sequence of mutually independent random variables $I_{X_1}(B), \ldots, I_{X_n}(B), \ldots$ obeys the strong law of large numbers, i.e. as $n \to \infty$, with probability 1

$$\frac{1}{n}\sum_{i=1}^{n} I_{X_i}(B) - \frac{1}{n}\sum_{i=1}^{n} EI_{X_i}(B) \to 0. \tag{2.20}$$

From (2.20) we have

$$P_n^*(B) - \frac{1}{n}\sum_{i=1}^{n} P_i(B) \to 0, \;\; B \in \Re$$

and

$$F_n^*(x) - \frac{1}{n}\sum_{i=1}^{n} F_i(x) \to 0, \;\; x \in \mathbb{R}.$$

Lemma 2.1 *For any $B \in \Re$ and $x \in \mathbb{R}$*

$$P\{nP_n^*(B) = k\} = \frac{1}{k!(n-k)!} \sum_{\wp_{1,2,\ldots,n}} \prod_{i=1}^{k} P_{j_i}(B) \prod_{i=k+1}^{n} (1 - P_{j_i}(B))$$

and

$$P\{nF_n^*(x) = k\} = \frac{1}{k!(n-k)!} \sum_{\wp_{1,2,\ldots,n}} \prod_{i=1}^{k} F_{j_i}(x) \prod_{i=k+1}^{n} (1 - F_{j_i}(x)),$$

where the summation $\wp_{1,2,\ldots,n}$ *extends over all* $n!$ *permutations* $(j_1, j_2, \ldots, j_n)$ *of* $(1, 2, \ldots, n)$.

Denote now

$$\mathbf{A}(n, k; x) = \binom{n}{k} x^k (1-x)^{n-k}, \tag{2.21}$$

$0 \le x \le 1,\ k = 0, 1, 2, \ldots, n;\ n \ge 1$ and the symmetric function

$$\mathbf{B}(n, k; x_1, x_2, \ldots, x_n) = \frac{1}{k!(n-k)!} \sum_{\wp_{1,2,\ldots,n}} \prod_{i=1}^{k} x_{j_i} \prod_{i=k+1}^{n} (1 - x_{j_i}), \tag{2.22}$$

$k = 0, 1, 2, \ldots, n;\ n \ge 1,\ 0 \le x_l \le 1,\ l = 1, 2, \ldots, n,$ where the summation $\wp_{1,2,\ldots,n}$ extends over all $n!$ permutations $(j_1, j_2, \ldots, j_n)$ of $(1, 2, \ldots, n)$ assuming $\prod_{i=j+1}^{j} a_i$ are equal to 1.

It is useful to note that

$$\mathbf{B}(n, n; x_1, x_2, \ldots, x_n) = x_1 x_2 \cdots x_n$$

and

$$\mathbf{B}(n, 0; x_1, x_2, \ldots, x_n) = (1 - x_1)(1 - x_2) \cdots (1 - x_n).$$

Since $\mathbf{B}(n, k; x_{j_1}, x_{j_2}, \ldots, x_{j_n}) = \mathbf{B}(n, k; x_1, x_2, \ldots, x_n)$ for all $n!$ permutations $(j_1, j_2, \ldots, j_n)$ of $(1, 2, \ldots, n)$, then

$$P\{nF_n^*(x) = k\} = \mathbf{B}(n, k; F_1(x), F_2(x), \ldots, F_n(x)). \tag{2.23}$$

If $F_1 = F_2 = \cdots = F_n = F$, then (2.23) becomes

$$P\{nF_n^*(x) = k\} = \mathbf{A}(n, k; F(x)). \tag{2.24}$$

The following recurrence relation can be useful.

Lemma 2.2 *For* $0 \leq x_l \leq 1,\ l = 1, 2, \ldots, n$, *the following recurrence relation is valid for* $k = 1, 2, \ldots, n-1$ *and* $n \geq 2$

$$\begin{aligned}\mathbf{B}(n, k; x_1, x_2, \ldots, x_n) &= \mathbf{B}(n-1, k; x_1, x_2, \ldots, x_{n-1})\bar{x}_n \\ &\quad + \mathbf{B}(n-1, k-1; x_1, x_2, \ldots, x_{n-1})x_n, \end{aligned} \tag{2.25}$$

where $\bar{x}_n = 1 - x_n$.

Proof. We can prove that assertion expressing the symmetric function $\mathbf{B}(n, k; x_1, x_2, \ldots, x_n)$ by permanent.

$$\begin{aligned}&B(n, k; x_1, x_2, \ldots, x_n) \\ &\quad = \frac{1}{k!(n-k)!} \sum_{\wp_{1,2,\ldots,n}} x_{j_1} x_{j_2} \cdots x_{j_k} \bar{x}_{j_{k+1}} \cdots \bar{x}_{j_n} \\ &\quad = \frac{1}{k!(n-k)!} Per\mathbf{A}_1 \end{aligned} \tag{2.26}$$

where $\bar{x}_i = 1 - x_i$, $i = 1, 2, \ldots, n$ and

$$Per\mathbf{A}_1 = \left\| \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ \cdots & \cdots & \cdots & \cdots \\ x_1 & x_2 & \cdots & x_n \\ & & & \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \\ \cdots & \cdots & \cdots & \cdots \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \end{bmatrix} \right\| \begin{matrix} \left.\vphantom{\begin{matrix} x \\ x \\ x \end{matrix}}\right\} k \text{ times} \\ \\ \left.\vphantom{\begin{matrix} x \\ x \\ x \end{matrix}}\right\} n-k \text{ times} \end{matrix}$$

It is easy to see that expanding the permanent along the last column we have

$$\left\| \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ \cdots & \cdots & \cdots & \cdots \\ x_1 & x_2 & \cdots & x_n \\ \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \\ \cdots & \cdots & \cdots & \cdots \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \end{bmatrix} \right\| \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} k \text{ times} \\ \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n-k \text{ times} \end{matrix}$$

$$= kx_n \left\| \begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ x_1 & x_2 & \cdots & x_{n-1} \\ \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ \bar{x}_1 & \bar{x}_2 & & \bar{x}_{n-1} \end{bmatrix} \right\| \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} k \ -1 \text{ times} \\ \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n \ -k \text{ times} \end{matrix} \tag{2.27}$$

$$+(n-k)\bar{x}_n \left\| \begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ x_1 & x_2 & \cdots & x_{n-1} \\ \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_{n-1} \end{bmatrix} \right\| \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} k \ \text{ times} \\ \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n-k-1 \text{ times} \end{matrix} \tag{2.28}$$

From (2.26) and (2.28), we obtain the assertion of the lemma. □

Using (2.25) the cdf of rth order statistic $X_{r:n}$ given in (2.7) can be shown in terms of symmetric functions as

$$F_{r:n}(x) = \sum_{i=r}^{n} \mathbf{B}(n, i, F_1(x), F_2(x), \ldots, F_n(x)). \tag{2.29}$$

and the relevant recurrence relationship may be expressed by using Lemma 2.2 as

$$F_{r:n}(x) = \bar{F}_n(x)F_{r:n-1}(x) + F_n(x)F_{r-1:n-1}(x),\ 1 \leq r < n. \qquad (2.30)$$

where $F_{r:n-1}$ denote the cdf of the rth order statistic from INID random variables $X_1, X_2, \ldots, X_{n-1}$ with corresponding cdf's $F_1, F_2, \ldots, F_{n-1}$.

The relation given in (2.30) may be obtained using Lemma 2.2

$$F_{r:n}(x) = \sum_{i=r}^{n} \mathbf{B}(n, i, F_1(x), F_2(x), \ldots, F_n(x))$$

$$= \sum_{i=r}^{n-1} \mathbf{B}(n, i, F_1(x), F_2(x), \ldots, F_n(x)) + \mathbf{B}(n, n, F_1(x), F_2(x), \ldots, F_n(x))$$

$$= \bar{F}_n(x) \sum_{i=r}^{n-1} \mathbf{B}(n-1, i, F_1(x), F_2(x), \ldots, F_{n-1}(x))$$

$$+ F_n(x) \sum_{i=r}^{n-1} \mathbf{B}(n-1, i-1, F_1(x), F_2(x), \ldots, F_{n-1}(x))$$

$$+ F_n(x)\mathbf{B}(n-1, n-1, F_1(x), F_2(x), \ldots, F_{n-1}(x))$$

$$= \bar{F}_n(x)F_{r:n-1}(x) + F_n(x) \sum_{j=r-1}^{n-2} \mathbf{B}(n-1, j, F_1(x), F_2(x), \ldots, F_{n-1}(x))$$

$$+ F_n(x)\mathbf{B}(n-1, n-1, F_1(x), F_2(x), \ldots, F_{n-1}(x)),$$

For $r = n$ we have $F_{n:n}(x) = F_{n-1:n-1}(x)F_n(x)$. Note that (2.30) and related recurrence equalities can be found in ([29], p. 105).

Since,

$$P\{nF_n^*(x) = i\} = \mathbf{B}(n, i, F_1(x), F_2(x), \ldots, F_n(x)), \quad i = 0, 1, 2, \ldots, n,$$

then

$$\sum_{i=0}^{n} \mathbf{B}(n, i, F_1(x), F_2(x), \ldots, F_n(x)) = 1. \tag{2.31}$$

We also have,

$$P\{X_{r:n} \leq x\} = \sum_{i=r}^{n} P\{nF_n^*(x) = i\}.$$

Now, for $0 \leq x_t < y_t \leq 1, t = 1, 2, \ldots, n$ denote by

$$\begin{aligned}
&\mathbf{C}(n, i, j; x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_n) \\
&\qquad = \frac{1}{i!(j-i)!(n-j)!} \sum_{\wp_{1,2,\ldots,n}} \prod_{l=1}^{i} x_{t_l} \\
&\qquad\quad \times \prod_{l=i+1}^{j} (y_{t_l} - x_{t_l}) \prod_{l=j+1}^{n} (1 - y_{t_l}),
\end{aligned} \tag{2.32}$$

where the summation $\wp_{1,2,\ldots,n}$ extends over all $n!$ permutations $(t_1, t_2, \ldots, t_n)$ of $(1, 2, \ldots, n)$.

Lemma 2.3 *Let $0 \leq x_t < y_t \leq 1, t = 1, 2, \ldots, n$. The following recurrence relation is valid for $1 \leq i < j < n$ and $n \geq 3$*

$$\begin{aligned}
&\mathbf{C}(n, i, j; x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_n) \\
&\quad = x_n \mathbf{C}(n-1, i-1, j-1; x_1, x_2, \ldots, x_{n-1}; y_1, y_2, \ldots, y_{n-1}) \\
&\qquad + (y_n - x_n) \mathbf{C}(n-1, i, j-1; x_1, x_2, \ldots, x_{n-1}; y_1, y_2, \ldots, y_{n-1}) \\
&\qquad + (1 - y_n) \mathbf{C}(n-1, i, j; x_1, x_2, \ldots, x_{n-1}; y_1, y_2, \ldots, y_{n-1}).
\end{aligned} \tag{2.33}$$

Proof. The proof can be made by mathematical induction or can be obtained from permanent expressions. $X_{r:n}$ and $X_{s:n}$

$$\mathbf{C}(n,i,j;x_1,x_2,\ldots,x_n;y_1,y_2,\ldots,y_n)=\frac{1}{i!(j-i)!(n-j)!}Per\mathbf{A}_2 \tag{2.34}$$

where

$$Per\mathbf{A}_2=\left|\begin{bmatrix} x_1 & \cdots & x_n \\ \vdots & & \vdots \\ x_1 & \cdots & x_n \\ \\ y_1-x_1 & \cdots & y_n-x_n \\ \vdots & & \vdots \\ y_1-x_1 & \cdots & y_n-x_n \\ \\ 1-y_1 & \cdots & 1-y_n \\ \vdots & & \vdots \\ 1-y_1 & \cdots & 1-y_n \end{bmatrix}\right| \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \text{i times} \\ \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \text{j-i times} \\ \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \text{n-j times} \end{matrix}$$

It is easy to expand $Per\mathbf{A}_2$ as follows

$$\left|\begin{bmatrix} x_1 & \cdots & x_n \\ \vdots & & \vdots \\ x_1 & \cdots & x_n \\ \\ y_1-x_1 & \cdots & y_n-x_n \\ \vdots & & \vdots \\ y_1-x_1 & \cdots & y_n-x_n \\ \\ 1-y_1 & \cdots & 1-y_n \\ \vdots & & \vdots \\ 1-y_1 & \cdots & 1-y_n \end{bmatrix}\right| \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \text{i times} \\ \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \text{j-i times} \\ \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \text{n-j times} \end{matrix}$$

$$= ix_n \left| \begin{bmatrix} x_1 & \cdots & x_{n-1} \\ \vdots & & \vdots \\ x_1 & \cdots & x_{n-1} \\ & & \\ y_1 - x_1 & \cdots & y_{n-1} - x_{n-1} \\ \vdots & & \vdots \\ y_1 - x_1 & \cdots & y_{n-1} - x_{n-1} \\ & & \\ 1 - y_1 & \cdots & 1 - y_{n-1} \\ \vdots & & \vdots \\ 1 - y_1 & \cdots & 1 - y_{n-1} \end{bmatrix} \right| \begin{matrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} & \text{i-1 times} \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} & \text{j-i times} \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} & \text{n-j times} \end{matrix}$$

$$+(j-i)(y_n - x_n) \left| \begin{bmatrix} x_1 & \cdots & x_{n-1} \\ \vdots & & \vdots \\ x_1 & \cdots & x_{n-1} \\ & & \\ y_1 - x_1 & \cdots & y_{n-1} - x_{n-1} \\ \vdots & & \vdots \\ y_1 - x_1 & \cdots & y_{n-1} - x_{n-1} \\ & & \\ 1 - y_1 & \cdots & 1 - y_{n-1} \\ \vdots & & \vdots \\ 1 - y_1 & \cdots & 1 - y_{n-1} \end{bmatrix} \right| \begin{matrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} & \text{i times} \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} & \text{j-i-1 times} \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} & \text{n-j times} \end{matrix}$$

$$+(n-j)(1-y_n)\left|\left[\begin{array}{ccc} x_1 & \cdots & x_{n-1} \\ \vdots & & \vdots \\ x_1 & \cdots & x_{n-1} \\ \\ y_1-x_1 & \cdots & y_{n-1}-x_{n-1} \\ \vdots & & \vdots \\ y_1-x_1 & \cdots & y_{n-1}-x_{n-1} \\ \\ 1-y_1 & \cdots & 1-y_{n-1} \\ \vdots & & \vdots \\ 1-y_1 & \cdots & 1-y_{n-1} \end{array}\right]\right| \begin{array}{l} \left.\vphantom{\begin{array}{c}x\\ \vdots\\ x\end{array}}\right\}\text{ i times} \\ \\ \left.\vphantom{\begin{array}{c}x\\ \vdots\\ x\end{array}}\right\}\text{ j-i times} \\ \\ \left.\vphantom{\begin{array}{c}x\\ \vdots\\ x\end{array}}\right\}\text{ n-j-1 times} \end{array} \qquad \square$$

Now we can be express the joint distribution function of order statistics $X_{r:n}$ and $X_{s:n}$ in terms of the symmetric function $\mathbf{C}(n,i,j;x_1,x_2,\ldots,x_n;y_1,y_2,\ldots,y_n)$ as follows:

$$\begin{aligned} F_{r,s}(x,y) &= \sum_{j=s}^{n}\sum_{i=r}^{j}\mathbf{C}(n,i,j;F_1(x),F_2(x),\ldots,F_n(x); \\ & \quad F_1(y),F_2(y),\ldots,F_n(y)). \end{aligned} \tag{2.35}$$

Using Lemma 2.3 one can write the following recurrence relation for $1 \le r < s \le n, n \ge 3$

$$\begin{aligned} F_{r,s:n}(x,y) &= F_n(x)F_{r-1,s-1:n-1}(x,y) + [F_n(y)-F_n(x)] \times \\ & \quad \times F_{r,s-1:n-1}(x,y) + (1-F_n(y))F_{r,s:n-1}(x,y), \end{aligned} \tag{2.36}$$

where $F_{r,s-1:n-1}(x,y)$ is the joint cdf rth and $(s-1)st$ order statistics from INID random variables $X_1, X_2, \ldots, X_{n-1}$ with corresponding cdf's $F_1, F_2, \ldots, F_{n-1}$. (see also [28], P.106).

Indeed, one can write

$$
\begin{aligned}
& F_{r,s}(x,y) \\
= \;& \sum_{i=r}^{n} \sum_{j=\max(0,s-i)}^{n-i} \mathbf{C}(n,i,j;F_1(x),F_2(x),\ldots,F_n(x);F_1(y),F_2(y),\ldots,F_n(y)) \\
= \;& \sum_{i=r}^{n} \sum_{j=\max(0,s-i)}^{n-i} [F_n(x)\mathbf{C}(n-1,i-1,j-1;F_1(x),F_2(x),\ldots,F_{n-1}(x);F_1(y); \\
& F_1(y),F_2(y),\ldots,F_{n-1}(y)) + (F_n(y)-F_n(x))\mathbf{C}(n-1,i,j-1; \\
& F_1(x),F_2(x),\ldots,F_{n-1}(x);F_1(y),F_2(y),\ldots,F_{n-1}(y)) \\
& +(1-F_n(y))B(n-1,i,j;F_1(x),F_2(x),\ldots,F_{n-1}(x);F_1(y),F_2(y),\ldots,F_{n-1}(y))] \\
= \;& \sum_{i=r}^{n-1} \sum_{j=\max(0,s-i)}^{n-i} F_n(x)\mathbf{C}(n-1,i-1,j-1;F_1(x),F_2(x),\ldots,F_{n-1}(x); \\
& F_1(y),F_2(y),\ldots,F_{n-1}(y)) \\
& +\sum_{i=r}^{n-1} \sum_{j=\max(0,s-i)}^{n-i} (F_n(y)-F_n(x))\mathbf{C}(n-1,i,j-1;F_1(x),F_2(x),\ldots,F_{n-1}(x); \\
& F_1(y),F_2(y),\ldots,F_{n-1}(y)) \\
& +\sum_{i=r}^{n-1} \sum_{j=\max(0,s-i)}^{n-i} (1-F_n(y))\mathbf{C}(n-1,i,j;F_1(x),F_2(x),\ldots,F_{n-1}(x); \\
& F_1(y),F_2(y),\ldots,F_{n-1}(y)) \\
= \;& F_n(x)F_{r-1,s-1:n-1}(x,y) + [F_n(y)-F_n(x)]F_{r,s-1:n-1}(x,y) \\
& +(1-F_n(y))F_{r,s:n-1}(x,y).
\end{aligned}
$$

Chapter 3

Exceedance Models

The exceedance statistics denote the total number of observations over a random threshold. Random thresholds were first used in the work of [38] and [31]. Exceedance models with random thresholds are generally constructed considering two independent random samples. After determining the random thresholds based on the first sample, one can investigate the behaviour of the observations in the second sample regarding the behaviour of these random thresholds. Random threshold values make the problem of finding the distribution of exceedance statistics complicated and difficult. However, if the threshold value is fixed, then the calculation is easier and closely related to the binomial model. A particular type of such a problem for the threshold being an order statistic from the initial sample is related to the concept of tolerance limits ([71]) and invariant confidence intervals containing the future observations ([16]). In the 1950's, the theory of tolerance limits were widely studied (see [38], [31], [74], [82]).

Exceedance statistics are also used for constructing a test for whether two monitored random samples are from the same population. Non-parametric tests of identity of distributions with respect to the exceedances have been demonstrated by [49], [50], [56], [44] and [47]. Another aspect of the theory of exceedances is being closely related to the inverse hypergeometric distribution (see [77], [38] and [74]).

3.1 Exceedances based on order statistics

3.1.1 Invariant confidence intervals

Invariant confidence intervals containing a future observation which is introduced by [16] is an important concept to mention in support of the use of order statistics as a random thresholds.

Definition 3.1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n with distribution function $F \in \mathfrak{F}$, where $\mathfrak{F}$ is some class of distribution functions. Assume that $g_1(.)$ and $g_2(.)$ are two Borel functions satisfying $g_1(x_1, x_2, \ldots, x_n) \leq g_2(x_1, x_2, \ldots, x_n)$, $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. The random interval $(g_1(X_1, X_2, \ldots, X_n),\ g_2(X_1, X_2, \ldots, X_n))$ is called an invariant confidence interval containing the future observations for class $\mathfrak{F}$, if $\exists \alpha \in (0, 1)$ such that

$$P\{X_{n+1} \in (g_1(X_1, X_2, \ldots, X_n), g_2(X_1, X_2, \ldots, X_n))\} = \alpha, \forall F \in \mathfrak{F}.$$

The quantity α is the same for all $F \in \mathfrak{F}$ and is called a confidence level of an invariant interval. Let $\mathfrak{F}_c$ be the class of all continuous distribution functions and $\mathfrak{F} = \mathfrak{F}_c$.

It is known from [16] that under some certain conditions any invariant confidence interval for the class $\mathfrak{F}_c$ can be constructed only by the order statistics. Therefore $g_1(X_1, X_2, \ldots, X_n) = X_{r:n}$, $g_2(X_1, X_2, \ldots, X_n) = X_{s:n}$, $1 \leq r < s \leq n$, where $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ are order statistics obtained from $X_1, X_2, \ldots, X_n$, which is defined with the probability 1, and

$$P\{X_{n+1} \in (X_{r:n}, X_{s:n})\} = \frac{s - r}{n + 1}, \tag{3.1}$$

i.e. $(X_{r:n}, X_{s:n})$ is a distribution free confidence interval containing a future observation in the class of all absolutely continuous distribution functions $\mathbb{F}_c$.

3.1.2 Exceedances from IID random variables

Suppose that $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size n from a population with an unknown absolutely continuous cumulative distribution function $F_X(.)$. Now suppose further observations X_{n+1}, $X_{n+2}, \ldots$, X_{n+N} are drawn from the same population, independent from $\mathbf{X}$. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics related to the first sample, where $1 \leq r \leq n$. It is easy to obtain the probability of a future observation being greater than the rth largest observation in the initial sample

$$P(X_{n+i} \in (X_{r:n}, \infty)) = \frac{n-r+1}{n+1}, \; i = 1, \ldots, N \tag{3.2}$$

Denote the number of future observations exceeding the rth largest value in the first sample by $S_r = \sum\limits_{i=1}^{N} I_{\{(X_{r:n},\infty)\}}(X_{n+i})$, where

$$I_{\{(X_{r:n},\infty)\}}(X_{n+i}) = \begin{cases} 1, & X_{n+i} \in (X_{r:n}, \infty) \\ 0, & otherwise \end{cases}$$

The distribution of the number of exceedances in N future observations over the rth largest value of the first sample is

$$P(S_r = x) = \frac{r\binom{n}{r}\binom{N}{x}}{(N+n)\binom{N+n-1}{m+x-1}}, \tag{3.3}$$

where $1 \leq r \leq n$, $0 \leq x \leq N$.
It can be obtained as shown below

$P\{S_r = x\}$

$$= \binom{N}{x} \int_{-\infty}^{\infty} P\{X_{n+1} \leq t, \ldots, X_{n+x} \leq t, X_{n+x+1} > t, \ldots, X_N > t\} f_{r:n}(t)dt$$

$$= \binom{N}{x}\binom{n}{r} r \int_{-\infty}^{\infty} F_X^{x+r-1}(t)(1 - F_X(t))^{N-x+n-r} dF_X(t)$$

$$= \binom{N}{x}\binom{n}{r} r Beta(x + r, N - x + n - r + 1)$$

where $Beta(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$.

Since the probability given in (3.3) depends on only the parameters n, m, and N, it is distribution-free and it is clear that

$$\sum_{i=0}^{N} P(S_r = x) = 1.$$

Further aspects of given distribution can be found in [38].

Assume that $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size n from a population with cumulative distribution function $F_X(\cdot)$. Consider another sample $\mathbf{Y} = Y_1, Y_2, \ldots, Y_N$ of size N with cumulative distribution function $F_Y(\cdot)$ and independent from $\mathbf{X}$. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be order statistics obtained from the first sample. It is clear that under the hypothesis $H_0 : F_X(\cdot) = F_Y(\cdot)$, (3.1) can be given

$$P\{Y_k \in (X_{r:n}, X_{s:n})\} = \frac{s - r}{n + 1},$$

where $1 \leq r < s \leq n, k = 1, 2, \ldots, N$.

Denote the number of future observations $Y_1, Y_2, \ldots, Y_N$ falling into interval $(X_{r:n}, X_{s:n})$ by $S_{rs} = \sum_{i=1}^{N} I_{\{(X_{r:n}, X_{s:n})\}}(Y_k)$, where

$$I_{\{(X_{r:n}, X_{s:n})\}}(Y_k) = \begin{cases} 1, & Y_k \in (X_{r:n}, X_{s:n}) \\ 0, & otherwise \end{cases}$$

The distribution of S_{rs} is

$$P\{S_{rs} = x\} = \frac{(s-r)\binom{N}{x}\binom{n}{s-r}}{(s-r+x)\binom{n+N}{x+s-r}}, x = 0, 1, 2, \ldots, N \tag{3.4}$$

(see theorem 2 in [16]) and also the joint pdf of future observations falling into $(X_{r:n}, X_{s:n})$

$$P\{X_{n+1},\, X_{n+2},\, \ldots,\, X_{n+N} \in (X_{r:n}, X_{s:n})\} = \frac{n!(N+s-r-1)!}{(s-r-1)!(N+n)!} \tag{3.5}$$

(see e.g., [16], [44], [56]).

For $s = n$ and $r = 1$, (3.5) can be easily written as

$$P\{X_{n+1},\, X_{n+2},\, \ldots,\, X_{n+N} \in (X_{1:n}, X_{n:n})\} = \frac{n(n-1)}{(n+N)!(N+n-1)!}.$$

When $H_0 : F_X(\cdot) = F_Y(\cdot) = F(\cdot)$, the asymptotic distribution of $\frac{S_{rs}}{N}$ is

$$\lim_{N\to\infty} \sup_{0\le x\le 1} \left| P\left\{\frac{S_{rs}}{N} \le x\right\} - P\{Z_{rs} \le x\}\right| = 0 \tag{3.6}$$

where $Z_{rs} = F_Y(X_{s:n}) - F_Y(X_{r:n})$ (see [3]). The probability density function of Z_{rs} given in (page 33, [28]) is

$$f(z_{rs}) = \begin{cases} \frac{1}{Beta(s-r,n-s+r+1)} z_{rs}^{s-r-1}(1-z_{rs})^{n-s+r} & \text{if} \quad 0 \le z_{rs} \le 1, \\ 0 & \quad\quad \text{otherwise} \end{cases},$$

that is, Z_{rs} has a $Beta(s-r, n-s+r+1)$ distribution. (3.6) can be extended for any random interval $(g_1(X_1,\ldots,X_n), g_2(X_1,\ldots,X_n))$ as

$$\lim_{N\to\infty} \sup_{0\le x\le 1} \left| P\left\{\frac{S_{rs}}{N} \le x\right\} - P\{F_Y(g_2(X_1,\ldots,X_n)) - F_Y(g_1(X_1,\ldots,X_n)) \le x\}\right|$$

$= 0$,
where $g_1(X_1,\ldots,X_n) \le g_2(X_1,\ldots,X_n)$ $\forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}$ (see [3] and [11]).

When $F_X(\cdot)$ is not necessarily equal to $F_Y(\cdot)$, the statement

$$\lim_{N\to\infty} \sup_{0\le x\le 1} \left| P\left\{\frac{S_{rs}}{N} \le x\right\} - P\{F_Y(X_{s:n}) - F_Y(X_{r:n})\} = 0 \right| = 0 \tag{3.7}$$

is true for any r and s satisfying $1 \le r < s \le n$.

3.2 Exceedance statistics from dependent random variables

In this section we present the distributional properties of exceedance statistics of finite Farlie-Gumbel-Morgenstern (FGM) sequences regarding a random threshold which is studied by [7]. FGM class of multivariate distributions was discussed by [57], [37] and [33] and improved specifically in the works of [45] and [46] who introduced an additional parameter to achieve stronger correlation structure.

3.2.1 FGM and s-FGM random variables

First we give the general definitions of random FGM and s-FGM sequences.

Definition 3.2. Let $\{X_i\}_{i\ge 1}$ be a random sequence of random variables with the cumulative distribution functions $F_1(x), F_2(x), \ldots, F_n(x)$, and $\alpha(\cdot,\cdot)$ be a symmetric function (i.e. $\alpha(m,k) = \alpha(k,m)$). If the joint distribution of $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ is

$$H_{i_1,i_2,\ldots,i_n}(x_1,\ldots,x_n) = \prod_{k=1}^{n} F_{i_k}(x_k) \left\{1 + \sum_{1\le m<k\le n} \alpha(i_m,i_k)\overline{F}_{i_m}(x_m)\overline{F}_{i_k}(x_k)\right\}, \tag{3.8}$$

where $\overline{F}(x) = 1 - F(x)$, then the sequence $\{X_i\}_{i\ge 1}$ is called *FGM random sequence.*

The following definition is important to understand the distributional behaviours of exceedances from dependent variates.

Definition 3.3. Let $(X_1, X_2, \ldots, X_n)$ be a n variate FGM random vector with the marginal cdfs $F_1(x), F_2(x), \ldots, F_n(x)$, respectively and α_n be a real number. If these random variables have the following joint distribution function

$$H_{1,2,\ldots,n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i) \left\{ 1 + \alpha_n \sum_{1 \le m < k \le n} \overline{F}_m(x_m)\overline{F}_k(x_k) \right\}, \tag{3.9}$$

where $\overline{F}_m(x) = 1 - F_m(x_m)$ and for a given n the real number α_n is admissible, that is

$$1 + \alpha_n \sum_{1 \le m < k \le n} \epsilon_m \epsilon_k \ge 0$$

holds for all $\epsilon_m = \pm 1$ then the random variables $(X_1, X_2, \ldots, X_n)$ are called *simple FGM random variables.*

For further discussions on s-FGM distribution, we refer the reader to [7].

3.2.2 Exceedances in multivariate s-FGM distributions

Suppose that X is a random variable with the continuous distribution function $F(\cdot)$ and $Y_1, Y_2, \ldots, Y_N$ are s-FGM random variables with identical marginal distribution and pdf $G(x_i)$ and pdf $g(x_i)$ $(-\infty \le x_i \le \infty$, $i = 1, \ldots, N)$, respectively. The joint pdf of $(Y1, Y2, \ldots, Y_N)$ is given as follows

$$h(x_1, \ldots, x_N) = \prod_{i=1}^{N} g(x_i) \left\{ 1 + \alpha_N \sum_{1 \le m < k \le N} (1 - 2G(x_m))(1 - 2G(x_k)) \right\}$$

with α_N satisfying

$$-\frac{1}{\binom{N}{2}} \le \alpha \le \frac{1}{[\frac{N}{2}]}, \tag{3.10}$$

where $[a]$ denotes the integer part of the real number a.

Denote the number of observations of $Y_1, Y_2, \ldots, Y_N$'s falling into interval

$(-\infty, X)$ by $S = \sum_{k=1}^{N} I_{\{(-\infty,X)\}}(Y_k)$, where

$$I_{\{(-\infty,X)\}}(Y_k) = \begin{cases} 1, & Y_k \in (-\infty, X) \\ 0, & otherwise \end{cases}$$

The exact distribution of S is obtained by [7] as given in the following theorem

Theorem 3.1 *For any integer $N \geq 1$ and real number α satisfying (3.10)*

$$\begin{aligned} P\{S = k\} \quad = \quad & \binom{N}{k} \Big[E(G^k(X)\overline{G}^{N-k}(X)) + \alpha_N \Big(\tfrac{k(k-1)}{2} E(G^k(X)\overline{G}^{N-k+2}(X)) \\ & -k(N-k)E(G^{k+1}(X)\overline{G}^{N-k+1}(X)) \\ & + \tfrac{(N-k)(N-k-1)}{2} E(G^{k+2}(X)\overline{G}^{N-k}(X)) \Big) \Big] \end{aligned} \tag{3.11}$$

where $k = 0, 1, \ldots, N$ and $\overline{G}(x) = 1 - G(x)$. (see theorem 4.1 in [7]).

Proof.

$$P\{S = k\}$$

$$= \binom{N}{k} P\{Y_{i_1} \leq X, \ldots, Y_{i_k} \leq X, Y_{i_{k+1}} > X, \ldots, Y_{i_N} > X\}$$

$$= \binom{N}{k} \int_{-\infty}^{\infty} P\{Y_{i_1} \leq x, \ldots, Y_{i_k} \leq x, Y_{i_{k+1}} > x, \ldots, Y_{i_N} > x\} f(x) dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \int_{x}^{\infty} \cdots \int_{x}^{\infty} h_{1,\ldots,k\ldots,N}(x_1, \ldots, x_N) dx_1 \ldots dx_N dF(x) \quad ,$$

$$= \binom{N}{k} \int_{-\infty}^{\infty} \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \int_{x}^{\infty} \cdots \int_{x}^{\infty} \Big[\prod_{i=1}^{N} g(x_i) \Big\{ 1 + \alpha_N \sum_{1 \leq m < k \leq N} (1 - 2G(x_m))$$

$$(1 - 2G(x_k)) \Big\} \Big] \times dx_1 \ldots dx_N dF(x)$$

By changing variable $G(x_i) = z_i$, $i = 1, \ldots, N$ we obtain the following integration

$$= \binom{N}{k} \int_{-\infty}^{\infty} \int_{0}^{G(x)} \cdots \int_{0}^{G(x)} \int_{G(x)}^{1} \cdots \int_{G(x)}^{1} \left[1 + \alpha_N \sum_{1 \le m < k \le N} (1 - 2z_m)(1 - 2z_k) \right]$$

$$du_1 \ldots dz_N dF(x)$$

$$= \binom{N}{k} \int_{-\infty}^{\infty} \Big[G^k(x) \overline{G}^{N-k}(x) + \alpha_N \Big(\frac{k(k-1)}{2} G^k(x) \overline{G}^{N-k+2}(x)$$

$$-k(N-k) G^{k+1}(x) \overline{G}^{N-k+1}(x)$$

$$+ \tfrac{(N-k)(N-k-1)}{2} G^{k+2}(x) \overline{G}^{N-k}(x) \Big] \, dF(x)$$

□

Corollary 3.2 *It is easy to obtain that under the hypothesis* $F(\cdot) = G(\cdot)$

$$P\{S = k\}$$

$$= \binom{N}{k} \Big[B(k+1, N-k+1) + \alpha_N \Big(\tfrac{k(k-1)}{2} B(k+1, N-k+3)$$

$$-k(N-k) B(k+2, N-k+2)$$

$$+ \tfrac{(m-k)(m-k-1)}{2} B(k+3, N-k+1) \ \Big) \Big]$$

$k = 0, 1, \ldots, N$ *and* $-\frac{1}{\binom{N}{2}} \le \alpha \le \frac{1}{\left[\frac{N}{2}\right]}$, *where* $B(a, b)$ *is a beta function.*

Assume that $X_1, X_2, \ldots, X_n$ is the finite s-FGM sequence of random variables and $Y_1, Y_2, \ldots, Y_N, \ldots$ is a sequence of IID random variables with distribution function F and G, respectively. Let $X_{1:n} \le X_{1:n} \le \cdots \le X_{1:n}$ be the order statistics corresponding to $X_1, X_2, \ldots, X_n$.

Denote the number of observations of $Y_1, Y_2, \ldots, Y_N$'s falling into interval

$(-\infty, X_{r:n})$ by $T = \sum_{i=1}^{N} I_{\{(-\infty, X_{r:n})\}}(Y_k)$, where

$$I_{\{(-\infty, X_{r:n})\}}(Y_k) = \begin{cases} 1, & Y_k \in (-\infty, X_{r:n}) \\ 0, & otherwise \end{cases}$$

The exact distribution of S is obtained by [7] as given in the following theorem

Theorem 3.3 *It is true that under the hypothesis $F(\cdot) = G(\cdot)$ for any integer $N \geq 1$ and $1 \leq r \leq n$*

$$P\{T = k\}$$

$$= \binom{N}{k} \sum_{i=r}^{n} (-1)^{i-r} \binom{i-1}{r-1} \binom{n}{i} \left[iB(i+k, N-k+1) \right. \tag{3.12}$$

$$\left. +\alpha_n \left(\frac{i^2(i-1)}{2} B(i+k, N-k+3) - i(i-1)B(i+k+1, N-k+2) \right) \right]$$

$k = 0, 1, \ldots, N$ and $-\frac{1}{\binom{N}{2}} \leq \alpha \leq \frac{1}{\left[\frac{N}{2}\right]}$, where $B(a, b)$ is a beta function.

Proof.

$$P\{T = k\}$$

$$= \binom{N}{k} P\{Y_{i_1} \leq X_{r:n}, \ldots, Y_{i_k} \leq X_{r:n}, Y_{i_{k+1}} > X_{r:n}, \ldots, Y_{i_N} > X_{r:n}\} \tag{3.13}$$

$$= \binom{N}{k} \int_{-\infty}^{\infty} P\{Y_{i_1} \leq x, \ldots, Y_{i_k} \leq x, Y_{i_{k+1}} > x, \ldots, Y_{i_N} > x\} f_{r:n}(x) dx$$

It is important to note that for symmetrically distributed random variables the distribution function of the rth largest order statistic can be given (page 99 in [29])

$$P(X_{r:n} \leq x) = \sum_{i=r}^{n} (-1)^{i-r} \binom{i-1}{r-1} \binom{n}{i} P(X_{i:i} \leq x) \tag{3.14}$$

Since for any s-variate marginals of $X_1,\ X_2, \ldots,\ X_n$ the α_i is equal to α_n,

$$P(X_{i:i} \le x) = F^i(x)\left\{1 + \alpha_n \frac{i(i-1)}{2}(1-F(x))^2\right\}.$$

Substituting (3.14) in (3.13), we obtain the following equation

$$P\{T=k\}$$

$$= \tbinom{N}{k} \int_{-\infty}^{\infty} G^k(x)(1-G(x))^{N-k} \sum_{i=r}^{n} (-1)^{i-r} \binom{i-1}{r-1}\binom{n}{i}$$

$$\times \left\{ iF^{i-1}(x)f(x) + \alpha_n \tfrac{i^2(i-1)}{2} F^{i-1}(x)(1-F(x))^2 f(x)\right.$$

$$\left. -\alpha_n i(i-1)F^i(x)(1-F(x))f(x)\right\} dx$$

Since $F(\cdot) = G(\cdot)$ and using the transformation $F(x) = z$, one can easily get

$$P\{T=k\}$$

$$= \tbinom{N}{k} \sum_{i=r}^{n} (-1)^{i-r} \tbinom{i-1}{r-1} \tbinom{n}{i} \left\{ i \int_0^1 z^{i+k-1}(1-z)^{N-k} dz \right.$$

$$\left. +\alpha_n \tfrac{i^2(i-1)}{2} \int_0^1 z^{i+k-1}(1-z)^{N-k+2} dz - \alpha_n i(i-1) \int_0^1 z^{i+k}(1-z)^{N-k+1} dz \right\}.$$

□

Corollary 3.4 *For $r = n$, (3.12) becomes*

$$P\{T=k\} = \tbinom{N}{k}[nB(n+k, N-k+1)$$

$$+\alpha_n \left(\tfrac{n^2(n-1)}{2} B(n+k, N-k+3) - n(n-1)B(n+k+1, N-k+2) \right)\Big]$$

3.3 Various examples of exceedance models

In this section we aim to create an understanding of the theory and applications of exceedances for the reader by discussing the findings and developments presented in several early and recent research papers.

3.3.1 Statistical prediction with respect to the tolerance limits

In the early 1940's, exceedance statistics are used by [81] in statistical quality control to make statistical predictions about measurements on a specific quality characteristic in the future production. [81] first supposed that a random sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of n observations of a certain characteristic are continuous random variables with common pdf $f(x)$.

Considering the order statistics $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ corresponding to the given sample, the problem of tolerance limits for one or two quality characteristics based on the initial sample is discussed. After a future sample $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$ of size m, the problems of prediction related to the second sample are considered in [81].

These problems can be briefly addressed as

1. What is the probability that at least m_0 of the values of $\mathbf{X}$ in the second sample will exceed the smallest observation in the initial sample?
2. What is the probability that at least m_0 of the values of $\mathbf{X}$ in the second sample will lie between the smallest and the largest observation in the initial sample?

Now it is interesting to see the findings about the given problems in the following sections.

General probability formula

Define $100S_\alpha\%$ tolerance limits $T_{lower}(x_1, x_2, \ldots, x_n)$ and $T_{upper}(x_1, x_2, \ldots, x_n)$ for probability level α as two functions of the $X's$ in the first sample. The probability that at least $100S_\alpha\%$ observations of the second indefinitely large sample will fall between T_{lower} and T_{upper} is α which is

$$P\left\{\int_{T_{lower}}^{T_{upper}} dF(x) \geq S_\alpha\right\} = \alpha. \tag{3.15}$$

When the size of the future sample is finite, the same notation can be used. However, this time we are interested in the largest integer m_α such that the probability that at least $100\overline{S}_\alpha\%$ ($\overline{S} = \frac{m_\alpha}{m}$) of the $Y's$ will lie between T_{lower} and T_{upper} is at least α.

Now considering the tolerance limits are the order statistics, the results might be simple and independent of the pdf $f(x)$. Remember $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ are the order statistics obtained from the initial sample. Assume that $r_1, r_2, \ldots, r_k$ are the integers such that $1 \leq r_1 < r_2 < \cdots < r_k \leq n$. Let $x_{r_1}, x_{r_2}, \ldots, x_{r_k}$ be real numbers. By integrating the product of the conditional probability that $m_1, m_2, \ldots, m_{k+1}$ of the values from second sample will fall in the intervals $(-\infty, x_{r_1}), (x_{r_1}, x_{r_2}), \ldots, (x_{r_k}, \infty)$ respectively given that the joint pdf of $X_{r_1:n}\, X_{r_2:n},\, \ldots,\, X_{r_k:n}$ and the joint pdf of $X_{r_1:n}\, X_{r_2:n},\, \ldots,\, X_{r_k:n}$ (given in (1.1))) with respect to the $x's$ we obtain

$$P\{m_0, m_1, \ldots, m_{k+1}\} =$$

$$\frac{m!n!(m_1 + r_1 - 1)!(m_2 + r_2 - r_1 - 1)! \cdots (m_k + r_k - r_{k-1} - 1)!(m_{k+1} + n - r_k)!}{(r_1 - 1)!(r_2 - r_1 - 1)! \cdots (r_k - r_{k-1} - 1)!(n - r_k)!(m + n)!m_1!m_2! \cdots m_{k+1}!} \tag{3.16}$$

One tolerance limit

Considering only one tolerance limit for instance the smallest value, $T_{lower} = X_{1:n}$, the probability that m_0 of the m future observations will exceed the smallest value of the first sample can be simply obtained using (3.16) for $k = 1$, $r_1 = 1$, $m_2 = m_0$ and $m_1 = m - m_0$

$$g_1(m_0) = \frac{nm!(m_0 + n - 1)}{m_0!(m + n)!} \tag{3.17}$$

Analogously, the probability that m_0 of the m future observations will exceed the largest value of the first sample can be simply obtained using (3.16) for $k = 1$, $r_1 = n$, $m_1 = m_0$ and $m_2 = m - m_0$.

The recurrence relation

$$g_1(m_0 - 1) = \frac{m_0}{m_0! + n + 1} g_1(m_0) \tag{3.18}$$

makes it easy to calculate the values of $g_1(m_0)$.

Two tolerance limits

Now considering two tolerance limits such that the smallest ($X_{1:n}$) and the largest ($X_{n:n}$) observations of the first sample, using (3.16) for $k = 2$, $r_1 = 1$, $r_2 = n$, $m_2 = m_0$, $m_3 = m - m_0 - m_1$ we obtain the joint distribution of m_0 and m_1

$$g(m_0, m_1) = n(n-1)\frac{m!(m_0 + n - 2)}{m_0!(m + n)!} \tag{3.19}$$

Summing up (3.19) with reference to m_1 from 0 to m_m_0, the probability that m_0 of the m future observations will lie between $X_{1:n}$ and $X_{n:n}$ is

$$g_2(m_0) = n(n-1)(m - m_0 + 1)\frac{m!(m_0 + n - 2)}{m_0!(m + n)!} \tag{3.20}$$

and the recurrence relation is

$$g_2(m_0 - 1) = \frac{m_0(m - m_0 + 2)}{m_0!(m - m_0 + 1)(m_0 + n - 2)} g_2(m_0).$$

The remainder of the paper [81] will focus on further discussions, such as tolerance limits for two quality characteristics and the problems

1. For given values of n, m and level of significance (α), what is the largest integer m_α such that the probability is at least α that $m_0 \geq m_\alpha$?
2. What is the limiting value of $S_\alpha = \frac{m_\alpha}{m}$ as m increases indefinitely (i.e. the second sample is the population)?

3.3.2 Some distribution free properties of statistics based on record values

Record values and exceedance statistics are of great importance in several research areas and real-life problems such as weather, economic and sports data. The statistical study of record values started with Chandler (1952). However it has now spread in different areas. As we mentioned before many theoretical studies based on exceedance statistics are also contributed to the studies of record statistics. For further details, we refer [1], [34], [61] and [64] to the reader. In this section we examine a couple of leading papers in this area which are [79] and [3].

The exceedance statistics in record models considered in [3] can be summarized as given in the following part.

Let $X_1, X_2, \ldots, X_n$ be a sequence of IID random variables with continuous cumulative distribution function (cdf) $F(\cdot)$. Suppose that the sequence $U(1), U(2), \ldots, U(n)$ is a sequence of upper record times defined as;

$$U(k) = \begin{cases} 1, & k = 1 \\ min\{j : j > U(n-1), X_j > X_{U(n-1)}\}, & k = n \end{cases} \tag{3.21}$$

where $k > 1$ and $X_{U(n)}$ denotes the nth upper record value. It's known that the distribution function of $X_{U(r)}$ (see [1]), for $r = 1, 2, \ldots, n$

$$F_r(x) = P\{X_{U(r) \leq x}\} = \frac{1}{(r-1)!} \int_{-\infty}^{x} \left(\ln \frac{1}{1-F(u)} \right)^{r-1} f(u)du, \; -\infty < x < \infty. \tag{3.22}$$

Hence the pdf of $X_{U(r)}$ may be easily seen as

$$f_r(x) = \frac{1}{(r-1)!} \left(\ln \frac{1}{1-F(x)} \right)^{r-1}, \; -\infty < x < \infty. \tag{3.23}$$

Suppose that $Y_1, Y_2, \ldots, Y_m$ is another sequence of IID random variables with continuous cumulative distribution function(cdf) $F(\cdot)$. Consider the record values $X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)}$ and $Y_1, Y_2, \ldots, Y_m$, for any $i = 1, 2, \ldots, m$ and $r = 1, 2, \ldots$ it is true that

$$P\{Y_i < X_{U(r)}\} = 1 - \frac{1}{2^r} \tag{3.24}$$

which can be proved as

$$\begin{aligned}
P\{Y_i \leq X_{U(r)}\} & \\
&= \int_{-\infty}^{\infty} P(Y_i < x) f_r(x)dx \\
&= \frac{1}{(r-1)!} \int_{-\infty}^{\infty} F(u) \left(\ln \frac{1}{1-F(u)} \right)^{r-1} dF(u) \\
&= \frac{1}{(r-1)!} \int_{0}^{1} u \left(\ln \frac{1}{1-u} \right)^{r-1} du \\
&= \frac{1}{(r-1)!} \int_{0}^{\infty} (z)^{r-1} (1-e^{-z})e^{-z}dz \\
&= 1 - \frac{1}{2^r}.
\end{aligned}$$

It's clear that for any $k, r = 1, 2, \ldots$ and $s > r$

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

Define $\xi_i(r) = 1$ if $Y_i < X_{U(r)}$ and $\xi_i(r) = 0$ otherwise for given r and denote the number of observations $Y_1, Y_2, \ldots, Y_m$ which are less then $X_{U(r)}$ by $S_m(r) = \sum_{i=1}^{m} \xi_i(r)$. It's important to note that since $\xi_1(r), \xi_2(r), \ldots, \xi_m(r)$ are generally dependent, we cannot consider Bernoulli trials here. The exact distribution of $S_m(r)$ is derived in [3] as In particular, for any $m = 1, 2, \ldots$ and $r = 1, 2, \ldots$

$$P\{S_m(r) = k\} = \frac{\binom{m}{k}}{(r-1)!} \int_0^\infty e^{-z(m-k+1)}(1 - e^{-z})^k z^{r-1} dz \tag{3.25}$$

where $k = 0, 1, \ldots, m$. Proof of (3.25) is given

$$\begin{aligned}
&P\{S_m(r) = k\} \\
&= \binom{m}{k} P\{Y_1 < X_{U(r)}, \ldots, Y_k < X_{U(r)}, Y_{k+1} > X_{U(r)}, \ldots, Y_m > X_{U(r)}\} \\
&= \binom{m}{k} \int_{-\infty}^{\infty} P\{Y_1 < z, \ldots, Y_k < z, Y_{k+1} > z, \ldots, Y_m > z\} f_r(z) dz \\
&= \binom{m}{k} \int_{-\infty}^{\infty} \underbrace{\int_{-\infty}^{z} \cdots \int_{-\infty}^{z}}_{\text{k}} \underbrace{\int_{z}^{\infty} \cdots \int_{z}^{\infty}}_{\text{m-k}} \frac{1}{(r-1)!} \left(\ln \frac{1}{1 - F(z)} \right)^{r-1} \\
&\quad \times dF(x_1) \cdots dF(x_m) dz
\end{aligned}$$

$$= \frac{\binom{m}{k}}{(r-1)!} \int_{-\infty}^{\infty} (1-F(z))^{m-k} F^k(z) \left(\ln \frac{1}{1-F(z)}\right)^{r-1} dF(z)$$

$$= \frac{\binom{m}{k}}{(r-1)!} \int_{0}^{1} (1-t)^{m-k} t^k \left(\ln \frac{1}{1-t}\right)^{r-1} dt$$

$$= \frac{\binom{m}{k}}{(r-1)!} \int_{0}^{\infty} e^{-u(m-k+1)} (1-e^{-u})^k u^{r-1} du$$

It's true that

$$\lim_{m\to\infty} \sup_{0\le x\le 1} \left| P\left\{\frac{S_m(r)}{m} \le x\right\} - \frac{1}{(r-1)!} \int_0^x \left(\ln \frac{1}{1-u}\right)^{r-1} du \right| = 0. \tag{3.26}$$

3.3.3 Models of exceedances based on records

[79] considered more general models of exceedances based on records of two independent sequences $\{X_n\}_{n\ge1}$ and $\{Y_n\}_{n\ge1}$ with distribution functions $F_X(\cdot)$ and $F_Y(\cdot)$, respectively. Investigating the behaviour of a sequence of IID random variables with respect to a random threshold, three exceedance statistics are introduced and their exact and asymptotic distributions are derived in [79]. Now we present the interesting theoretical results of this paper.

Let X be a random variable with a distribution function $F_X(\cdot)$ and $\mathbf{Y} = \{Y_1, Y_2, \ldots, Y_m\}$ be a sequence of IID random variables with a common distribution $F_Y(\cdot)$, independent of X. Let $Y_{1:m}, \ldots, Y_{m:m}$ be the corresponding order statistics. Denote the number of observations in the sample $\mathbf{Y}$ which is below the random threshold X by

$$S_r = \#\{r \le m : Y_r \le X\},$$

for $m \ge 1$. It can also be defined as follows

$$S_r = \max\{r \le m : Y_{r:m} \le X\}.$$

For any integer $m \geq 1$, the distribution, mean and variance of S_m are given as

$$P\{S_r = x\} = \binom{m}{x} E(F_Y^x(X)\overline{F}_Y^{m-x}(X)),\ x = 0, 1, \ldots, m,$$

$E(S_r) = mE(F_Y(X))$ and $Var(S_r) = mE(F_Y(X)\overline{F}_Y(X)) + m^2 Var(F_Y(X))$, where $\overline{F}_Y(X)) = 1 - F_Y(X)$.

Secondly, the number T_r of Y's below the random level X in a sample of the size $T_r + m$ with the last observation exceeding the level X is considered.

For any integer $m \geq 1$

$$T_r = \#\{r \geq 0 : S_{m+r-1} = r, Y_{m+r} > X\}$$

In terms of order statistics

$$T_r = \min\{r \geq 0 : Y_{r+1:m+r} > X\}$$

The distribution and the first two conditional moments of T_r can be given as follows.

Assume that $P(F_Y(X) < 1) > 0$. For any integer $m \geq 1$

$$P(T_r = x) = \binom{m+x-1}{m-1} E(F_Y^x(X)\overline{F}_Y^m(X)), x = 0, 1, \ldots$$

$E(T_r) = mE(\frac{F_Y(X)}{\overline{F}_Y(X)} I_{(0,1)}(F_Y(X)))$ and

$Var(T_r) = mE(\frac{F_Y(X)}{\overline{F}_Y^2(X)} I_{(0,1)}(F_Y(X))) + m^2 Var(\frac{F_Y(X)}{\overline{F}_Y(X)} I_{(0,1)}(F_Y(X)))$

The last exceedance investigated in [79] was the number of records K of the sequence $\mathbf{Y}$ falling below the random threshold X. Let $U(r)$ be the rth record time for the sequence $\mathbf{Y}$, $r = 1, 2, \ldots$, i.e. $U(1) = 1$ and $U(r) = \min\{i > U(r-1) : Y_i > Y_{U(r-1)}\}$, $r = 2, 3, \ldots$ Then

$$K = \min\{r \geq 0 : Y_{U(r+1)} > X\}.$$

Assume that $P(F_Y(X) < 1) > 0$. The exact distribution, mean and variance of K is

$$P(K = x) = \frac{1}{x!} E(\overline{F}_Y(X)(-\log(\overline{F}_Y(X)))^r I_{(0,1)}(F_Y(X))), x = 0, 1, \ldots$$

$E(K) = -E(\log(\overline{F}_Y(X)) I_{(0,1)} F_Y(X))$ and

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

Chapter 4

Main Results

Assume that $X_1,\ X_2,\ \ldots,\ X_n$ are independent random variables with continuous distribution functions $F_1(t), F_2(t), \ldots, F_n(t)$, respectively and $Y_1,\ Y_2,\ \ldots,\ Y_m$ are independent copies of random variable Y with continuous distribution function G. Let $X_{1:n} \leq \cdots \leq X_{n:n}$ be the order statistics constructed from $X_1,\ X_2,\ \ldots,\ X_n$. For $1 \leq r < s \leq n$, let us define the random variable

$$S_m = \sum_{i=1}^{m} \xi_i,$$

where ξ_i is zero when $Y_{i_1} \in (X_{r:n}, X_{s:n})$ and 1 otherwise. It is clear that S_m is the number of observations falling into random threshold $(X_{r:n}, X_{s:n})$ and is the exceedance statistic defined as

$$\begin{aligned} & P\{S_m = k\} \\ &= \sum_{i_1,i_2,\ldots,i_m} P\{\xi_{i_1} = 1, \ldots, \xi_{i_k} = 1, \xi_{i_{k+1}} = 0, \ldots, \xi_{i_m} = 0\} \\ &= \sum_{i_1,i_2,\ldots,i_m} P\{Y_{i_1} \in (X_{r:n}, X_{s:n}), \ldots, Y_{i_k} \in (X_{r:n}, X_{s:n}), \qquad (4.1) \\ & Y_{i_{k+1}} \notin (X_{r:n}, X_{s:n}), \ldots, Y_{i_m} \notin (X_{r:n}, X_{s:n})\} \end{aligned}$$

The distribution theory of exceedance statistics have been studied in numerous papers which appeared in recent years in statistical literature. See e.g. [3], [79], [5], [6], [7], [13], [12] and [4].

In general, the derivation of the distribution of exceedance statistic S_m faces technical difficulties connected with the permanent expressions for joint distribution function. Indeed, one has

$$P\{S_m = k\} = \sum_{i_1,i_2,\dots,i_m} \int\int P\{Y_{i_1} \in (x,y),\dots,Y_{i_k} \in (x,y), \quad (4.2)$$

$$Y_{i_{k+1}} \notin (x,y),\dots,Y_{i_m} \notin (x,y)\} f_{r,s}(x,y)dxdy,$$

where the summation extends over all $m!$ permutations $(i_1,i_2,\dots,i_m)$ of $1,2,\dots,m$.

Considering the formula (1.4) one can observe that even for the special distributions $F_1,F_2,\dots,F_n$ the calculation of $P\{S_m = k\}$ meets with great difficulties and the formula that can be obtained is not convincing for applications. However, the asymptotic distribution of $\frac{S_m}{m}$ can be found by using the functional representations using empirical distribution functions.

In this study we focus on asymptotic distributions of exceedance statistics based on INID random variables. We show that $\frac{S_m}{m}$ converges in distribution to the random variable $G(X_{s:n}) - G(X_{r:n})$. Afterwards, we investigate some special distributions for which the distribution of exceedance statistics can be expressed in a good form. More precisely, we consider the F^α scheme introduced by [62] (see also [62], [67], [68]) and in a special case when $r = 1$ and $s = n$ derive the distribution function of $G(X_{n:n}) - G(X_{1:n})$.

Theorem 4.1 *Assume that $X_1,X_2,\dots,X_n,\dots$ and $Y_1,Y_2,\dots,Y_n,\dots$ are idependent. It is true that*

$$\lim_{m\to\infty} \sup_{0\le x\le 1} \left\{ \left| P\left\{\frac{S_m}{m} \le x\right\} - P\left\{W_{rs} \le x\right\} \right| \right\} = 0, \quad (4.3)$$

where $W_{rs} = G(X_{s:n}) - G(X_{r:n})$.

Proof. We have

$$S_m = \sum_{i=1}^{m} \xi_i = \sum_{i=1}^{m} I_{\{(X_{r:n}\ \ X_{s:n})\}}(Y_i) \tag{4.4}$$

where $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 1$ if $x \notin A$. Using the representation (4.4) and conditioning on $X_{r:n}$ and $X_{s:n}$ we can write

$$\begin{aligned}
& P\left\{\tfrac{S_m}{m} \leq x\right\} \\
= \ & P\left\{\frac{1}{m}\sum_{i=1}^{m} I_{\{(X_{r:n},X_{s:n})\}}(Y_i) \leq x\right\} \\
= \ & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\frac{1}{m}\sum_{i=1}^{m} I_{\{(X_{r:n},X_{s:n})\}}(Y_i) \leq x | X_{r:n} = t, X_{s:n} = z\right\} dF_{X_{r:n},X_{s:n}}(t,z) \\
= \ & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\frac{1}{m}\sum_{i=1}^{m} I_{\{(t,z)\}}(Y_i) \leq x\right\} dF_{X_{r:n},X_{s:n}}(t,z) \qquad (4.5) \\
= \ & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty} I_{\{(t,z)\}}(u)dG_m^*(u) \leq x\right\} dF_{X_{r:n},X_{s:n}}(t,z)
\end{aligned}$$

where $G_m^*(u) = \frac{1}{m}\sum_{i=1}^{m} I_{\{Y_i \leq u\}}$is the empirical distribution function of sample $Y_1, Y_2, \ldots, Y_m$.

Denote the functional of G as

$$\Im(G) = \int_{-\infty}^{\infty} I_{(t,z)}(u)dG(u) \tag{4.6}$$

and then

$$\Im(G_m^*) = \int_{-\infty}^{\infty} I_{(t,z)}(u)dG_m^*(u).$$

Since the functional $\Im(.)$ is continuous according to uniform metric, and using

Glivenko-Cantelli Theorem $P\{w : \sup_u |G^*_m(u) - G(u)| \to 0\} = 1$ we have

$$\Im(G^*_m) \to \Im(G), \text{ a.s. as } m \to \infty,$$

i.e.

$$P\left\{w : \lim_{m\to\infty} \Im(G^*_m) = \Im(G)\right\} = 1.$$

Then from (4.5) we have for $m \to \infty$

$$\begin{aligned}
& P\left\{\tfrac{S_m}{m} \le x\right\} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty} I_{(t,z)}(u)dG^*_m(u) \le x\right\} dF_{X_{r:n},X_{s:n}}(t,z) \\
&\quad \to \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty} I_{(t,z)}(u)dG(u) \le x\right\} dF_{X_{r:n},X_{s:n}}(t,z) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left\{\int_{-\infty}^{\infty} I_{(t,z)}(u)dG(u) \le x \mid X_{r:n} = t, X_{s:n} = z\right\} dF_{X_{r:n},X_{s:n}}(t,z) \\
&= P\left\{\int_{-\infty}^{\infty} I_{(X_{r:n},X_{s:n})\}}(u)dG(u) \le x\right\} \\
&= P\left\{\int_{X_{r:n}}^{X_{s:n}} dG(u) \le x\right\} \\
&= P\left\{G(X_{s:n}) - G(X_{r:n}) \le x\right\}
\end{aligned}$$

Thus the Theorem is proved. □

Remark 4.2 *The distribution function of the random variable W_{rs} in case of independent and identically distributed random variables can be found in [28].*

For INID random variables the distribution of W_{rs} in general has complicated form. However, for some special cases this distribution can be easily calculated.

In the following theorem the pdf of $W_{1n} = G(X_{n:n}) - G(X_{1:n})$ is derived. It is interesting that this pdf can be expressed in terms of permutations with inversions, incomplete beta function and hypergeometric functions. For permutation with inversions see [51] and for incomplete beta function and hypergeometric functions see [22]. Below we provide information about permutations with inversions which can be found e.g. [55].

4.1 Asymptotic distribution of the exceedance statistics from INID random variables

Permutation Inversions

Definition 4.1. Let $a_1, a_2, \ldots, a_n$ be a permutation of the set $\{1, 2, \ldots, n\}$. If $i < j$ and $a_i > a_j$, the pair (a_i, a_j) is called an "inversion" of the permutation.

For example, the permutation 4312 has five inversions which are $(4, 3)$, $(4, 1)$, $(4, 2)$, $(3, 1)$ and $(3, 2)$. Each inversion is a pair of elements that is *out of order*, and it's clear that the only permutation with no inversions is the unordered permutation. Let $I_n(k)$ denote the number of permutations of length n with k inversions. In the following an explicit formula for $I_n(k)$ when $k \leq n$ (see [51]) is given

$$\begin{aligned} I_n(k) &= \binom{n+k-1}{k} + \sum_{j=1}^{\infty} (-1)^j \binom{n+k-u_j-j-1}{k-u_j-j} \\ &\quad + \sum_{j=1}^{\infty} (-1)^j \binom{n+k-u_j-1}{k-u_j}. \end{aligned}$$

The binomial coefficients are defined to be zero when the lower index is negative. The u_j are the pentagonal numbers defined as

$$u_j = \frac{j(3j-1)}{2} \qquad j = 1, 2, \ldots .$$

$I_n(k) = I_n(\binom{n}{2} - k)$											
	k, number of inversions										
$n\backslash k$	0	1	2	3	4	5	6	7	8	9	10	11
1	**1**											
2	1	**1**										
3	1	2	**2**	1								
4	1	3	5	**6**	5	3	1					
5	1	4	9	15	**20**	22	20	15	9	4	1	
6	1	5	14	29	49	**71**	90	101	101	90	71	49
7	1	6	20	49	98	169	**259**	359	455	531	573	573
8	1	7	27	76	174	343	602	**961**	1415	1940	2493	3017
9	1	8	35	111	285	628	1230	2191	**3606**	5545	8031	11021
10	1	9	44	155	440	1068	2298	4489	8095	**13640**	21670	32683

Table 1. The exact value of $I_n(k)$ for specific n and k values

Theorem 4.3 *Let* $X_1, X_2, \ldots, X_n, \ldots$ *be a sequence of independent random variables with the continuous distribution functions* $F_1, F_2, \ldots, F_n, \ldots,$ *respectively and* $Y_1,\ Y_2,\ \ldots,\ Y_m$ *be independent copies of random variable* Y *with continuous distribution function* G *where*

$$F_i(t) = G^i(t), \quad \forall t \in R, \quad i = 1, 2, \ldots, n. \tag{4.7}$$

Consider $W_n \equiv W_{1n} = G(X_{n:n}) - G(X_{1:n})$. *Then the pdf of* W_{1n} *is*

$$
\begin{aligned}
& f_{W_n}(w) \\
& = I_n(0)\left[w^{n-2}\frac{2\binom{n}{2}}{\left(\binom{n}{2}+1\right)}\left(1-w^{\binom{n}{2}+1}\right)+nw^{n-1}\left(1-w^{\binom{n}{2}}\right)\right] \\
& +\sum_{i=1}^{\binom{n}{2}} I_n(i)\left\{2(-1)^{-(i+1)}\binom{n}{2}w^{\binom{n+1}{2}-1}Beta\left[1-\tfrac{1}{w},i+1,\binom{n}{2}-i+1\right]\right. \\
& +n(-1)^{-(i+1)}\left(\binom{n}{2}-i\right)w^{\binom{n+1}{2}}Beta\left[1-\tfrac{1}{w},i+1,\binom{n}{2}-i\right] \\
& -n\left(1-w\right)^{i}w^{n-2}\ {}_2F_1\left[1,\binom{n}{2}+1,i+1,1-\tfrac{1}{w}\right] \\
& \left.+\ i(-1)^{-(i+1)}\left(\binom{n}{2}-i\right)w^{\binom{n+1}{2}-1}Beta\left[1-\tfrac{1}{w},i,\binom{n}{2}-1\right]\right\}
\end{aligned}
\tag{4.8}
$$

if $0 \leq w \leq 1$ *and* $f_W(w) = 0$, *otherwise. In (4.8)* ${}_2F_1\left[a,b,c,z\right]$ *is the Gaussian hypergeometric function which is defined as*

$$
\begin{aligned}
{}_2F_1\left[a,b,c,z\right] &= 1+\tfrac{ab}{1!c}z+\tfrac{a(a+1)b(b+1)}{2!c(c+1)}z^2+\cdots \\
&= \sum_{n=0}^{\infty}\tfrac{(a)_n(b)_n}{(c)_n}\tfrac{z^n}{n!}
\end{aligned}
$$

and $(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)} = x(x+1)\ldots(x+n-1)$ *for* $n \geq 0$, *and* $Beta[z,a,b]$ *is the incomplete beta function defined by*

$$
Beta[z,a,b] \equiv \int_0^z u^{a-1}(1-u)^{b-1}du.
$$

Remark 4.4 *In special cases when* $n = 2, n = 3$,*the pdf given in (4.8) is*

$$
\begin{aligned}
f_{W_2}(w) &= 2-2w,\ 0 \leq w \leq 1 \\
f_{W_3}(w) &= 9w-18w^2+14w^3-5w^4,\ 0 \leq w \leq 1,
\end{aligned}
$$

respectively. It is clear that these densities are polynomial for any n.

Before proving Theorem 4.3 we need some auxiliary lemmas. Also the following theorem due to [25] will be used in the proof.

Theorem 4.5 *([25]. see [27] Section 6.4, Theorem B) The number $I_n(k)$ of permutations of n with k inversions satisfies the following recurrence relations:*

$$I_n(k) = \sum_{\max(0,k-n+1)\leq j\leq k} I_{n-1}(j)$$

for each $n \geq 1$. $I_n(0) = 1$ for each $n \geq 1$ and $I_0(k) = 1$ for each $k \geq 1$.

Lemma 4.6 *For any $x, y \in \mathbb{R}$ and positive integer n the following identity is valid*

$$\prod_{i=1}^{n}(y^i - x^i) = (y-x)^n \sum_{i=0}^{\binom{n}{2}} I_n(i) y^{\binom{n}{2}-i} x^i.$$

Proof. We use mathematical induction. For $n = 1$ we have

$$\begin{aligned}\prod_{i=1}^{1}(y^i - x^i) &= (y-x) = (y-x)\sum_{i=0}^{\binom{1}{2}} I_1(i) y^{\binom{1}{2}-i} x^i \\ &= (y-x)\left[I_1(0)y^0x^0\right] = (y-x),\end{aligned}$$

and for $n = 2$ we have

$$\begin{aligned}\prod_{i=1}^{2}(y^i - x^i) &= (y-x)(y^2-x^2) = (y-x)^2\sum_{i=0}^{\binom{2}{2}} I_2(i) y^{\binom{2}{2}-i} x^i \\ &= (y-x)^2\left[I_2(0)y^1x^0 + I_2(1)y^0x^1\right] \\ &= (y-x)^2(x+y) = (y-x)(y^2-x^2).\end{aligned}$$

Therefore the assertion of the lemma is clearly true for $n = 1$ and $n = 2$.

Now, using mathematical induction we will show that if for each $n \geq 1$ it is true that

$$\prod_{i=1}^{n}(y^i - x^i) = (y-x)^n \sum_{i=0}^{\binom{n}{2}} I_n(i) y^{\binom{n}{2}-i} x^i,$$

then

$$\prod_{i=1}^{n+1}(y^i - x^i) = (y-x)^{n+1} \sum_{i=0}^{\binom{n+1}{2}} I_{n+1}(i) y^{\binom{n+1}{2}-i} x^i.$$

Indeed, one has

$$\prod_{i=1}^{n+1}(y^i - x^i) = (y^{n+1} - x^{n+1}) \prod_{i=1}^{n}(y^i - x^i)$$

$$= (y-x)(y^n + y^{n-1}x + \cdots + yx^{n-1} + x^n) \prod_{i=1}^{n}(y^i - x^i)$$

$$= (y-x)^{n+1}(y^n + y^{n-1}x + \cdots + yx^{n-1} + x^n) \sum_{i=0}^{\binom{n}{2}} I_n(i) y^{\binom{n}{2}-i} x^i$$

$$= (y-x)^{n+1} \sum_{k=0}^{n} \sum_{i=0}^{\binom{n}{2}} I_n(i) y^{\binom{n}{2}-i+n-k} x^{i+k}$$

$$= (y-x)^{n+1} \sum_{k=0}^{n} \Big[I_n(0) y^{\binom{n}{2}+n-k} x^k + I_n(1) y^{\binom{n}{2}+n-k-1} x^{k+1} + \cdots$$

$$+ I_n(\tbinom{n}{2} - 1) y^{n+k+1} x^{\binom{n}{2}+k-1} + I_n(\tbinom{n}{2}) x^{\binom{n}{2}+k} \Big].$$

Therefore,

$$\prod_{i=1}^{n+1}(y^i - x^i) = (y^{n+1} - x^{n+1})$$

$$\times \Big[I_n(0)y^{\binom{n}{2}+n} + I_n(1)y^{\binom{n}{2}+n-1}x + \cdots + I_n(\tbinom{n}{2})y^n x^{\binom{n}{2}}$$

$$+ I_n(0)y^{\binom{n}{2}+n-1}x + I_n(1)y^{\binom{n}{2}+n-2}x^2 + \cdots + I_n(\tbinom{n}{2})y^{n-1}x^{\binom{n}{2}+1}$$

$$+ \cdots$$

$$+ I_n(0)y^{\binom{n}{2}+1}x^{n-1} + I_n(1)y^{\binom{n}{2}}x^n + \cdots + I_n(\tbinom{n}{2})yx^{\binom{n}{2}+n-1}$$

$$+ I_n(0)y^{\binom{n}{2}}x^n + I_n(1)y^{\binom{n}{2}-1}x^{n+1} + \cdots + I_n(\tbinom{n}{2})x^{\binom{n}{2}+n} \Big]$$

and

$$\begin{aligned}
\prod_{i=1}^{n+1}(y^i - x^i) &= (y^{n+1} - x^{n+1}) \Big[I_n(0)y^{\binom{n}{2}+n} \\
&+ (I_n(0) + I_n(1))\, y^{\binom{n}{2}+n-1}x \\
&+ (I_n(0) + I_n(1) + I_n(2))\, y^{\binom{n}{2}+n-2}x^2 + \cdots \\
&+ (I_n(0) + I_n(1) + \cdots + I_n(n))\, y^{\binom{n}{2}}x^n \\
&+ (I_n(1) + I_n(2) + \cdots + I_n(n+1))\, y^{\binom{n}{2}-1}x^{n+1} \\
&+ (I_n(2) + I_n(3) + \cdots + I_n(n+2))\, y^{\binom{n}{2}-2}x^{n+2} + \cdots \\
&+ I_n(\tbinom{n}{2} + n)x^{\binom{n}{2}+n} \Big].
\end{aligned}$$

Finally, using Theorem 4.5 one can write

$$\prod_{i=1}^{n+1}(y^i - x^i) = (y^{n+1} - x^{n+1})$$

$$\Bigg[\sum_{\max(0,0-(n+1)+1)\leq j\leq 0} I_n(j) \qquad y^{\binom{n}{2}+n} + \sum_{\max(0,1-(n+1)+1)\leq j\leq 1} I_n(j) \qquad y^{\binom{n}{2}+n-1}x + \cdots$$

$$+ \sum_{\max(0,\binom{n}{2}+n-(n+1)+1)\leq j\leq \binom{n}{2}+n} I_n(j) \qquad x^{\binom{n}{2}+n}\Bigg]$$

$$= (y^{n+1} - x^{n+1})\left[I_{n+1}(0)y^{\binom{n+1}{2}} + I_{n+1}(1)y^{\binom{n+1}{2}-1}x + \cdots\right.$$

$$\left. + I_{n+1}(\tbinom{n+1}{2})x^{\binom{n+1}{2}}\right]$$

$$= (y-x)^{n+1}\sum_{i=0}^{\binom{n+1}{2}} I_{n+1}(i)y^{\binom{n+1}{2}-i}x^i,$$

which proves the Lemma. □

Lemma 4.7 *Let the conditions of the Theorem 4.3 be satisfied. Denote by $H(x, y)$ the joint distribution function of $G(X_{1:n})$ and $G(X_{n:n})$. It is true that*

$$H(x, y) = y^{\frac{n(n+1)}{2}} - (y-x)^n \sum_{i=0}^{\binom{n}{2}} I_n(i)y^{\binom{n}{2}-i}x^i, \tag{4.9}$$

for $0 \leq x \leq y$ and $0 \leq y \leq 1$, where $I_n(k)$ denote the number of permutations of length n with k inversions.

Proof.

$$
\begin{aligned}
&H(x,y)\\
&= P\left\{G(X_{1:n}) \le x, G(X_{n:n}) \le y\right\}\\
&= P\left\{X_{n:n} \le G^{-1}(y)\right\} - P\left\{G^{-1}(x) < X_{1:n}, X_{n:n} \le G^{-1}(y)\right\}\\
&= P\left\{X_1 \le G^{-1}(y), \ldots, X_n \le G^{-1}(y)\right\}\\
&-P\left\{G^{-1}(x) < X_1 \le G^{-1}(y), \ldots, G^{-1}(x) < X_n \le G^{-1}(y)\right\}\\
&= \prod_{i=1}^{n} F_i(G^{-1}(y)) - \prod_{i=1}^{n} \left[F_i(G^{-1}(y)) - F_i(G^{-1}(x))\right]\\
&= \prod_{i=1}^{n} y^i - \prod_{i=1}^{n} \left[y^i - x^i\right],
\end{aligned}
\tag{4.10}
$$

where $G^{-1}(x) = \min\{t \in \mathbb{R} : G(t) \ge x\}$ is the quantile function of G.

Then using Lemma 4.6 the last expression given in (4.10) can be written in terms of permutation with inversions as

$$H(x,y) = y^{\frac{n(n+1)}{2}} - (y-x)^n \sum_{i=0}^{\binom{n}{2}} I_n(i) y^{\binom{n}{2}-i} x^i.$$

□

Remark 4.8 *[59] showed that*

$$\frac{1}{(1-t)^n} \prod_{i=1}^{n} (1-t^i) = \sum_{i=0}^{\binom{n}{2}} I_n(i) t^i. \tag{4.11}$$

Using (4.11) one has

$$\begin{aligned}\prod_{i=1}^{n}(y^i - x^i) &= y^{\frac{n(n+1)}{2}}\prod_{i=1}^{n}\left(1-\left(\frac{x}{y}\right)^i\right)\\ &= y^{\frac{n(n+1)}{2}}\left(1-\left(\frac{x}{y}\right)\right)^n\sum_{i=0}^{\binom{n}{2}} I_n(i)\left(\frac{x}{y}\right)^i\end{aligned}$$

and

$$\begin{aligned}H(x,y) &= \prod_{i=1}^{n} y^i - \prod_{i=1}^{n}(y^i - x^i)\\ &= y^{\frac{n(n+1)}{2}}\left\{1-\left(1-\left(\frac{x}{y}\right)\right)^n\sum_{i=0}^{\binom{n}{2}} I_n(i)\left(\frac{x}{y}\right)^i\right\}. \qquad (4.12)\end{aligned}$$

Corollary 4.9 *Let the function $h(x,y)$ denote the joint probability density function of $G(X_{1:n})$ and $G(X_{n:n})$. Then*

$$\begin{aligned}h(x,y) &= (y-x)^n\sum_{i=0}^{\binom{n}{2}} I_n(i)y^{\binom{n}{2}-i}x^i\\ &\times\left[\frac{2\binom{n}{2}}{(y-x)^2}+\frac{n\left(\binom{n}{2}-i\right)}{(y-x)y}-\frac{ni}{(y-x)x}-\frac{\left(\binom{n}{2}-i\right)i}{yx}\right] \qquad (4.13)\end{aligned}$$

for $0 \le x \le y$ and $0 \le y \le 1$, where $I_n(k)$ denote the number of permutations of length n with k inversions.

Proof.

$$h(x,y) = \frac{\partial^2 H(x,y)}{\partial x \partial y}$$

$$= -\frac{\partial^2 (y-x)^n \sum_{i=0}^{\binom{n}{2}} \left[I_n(i) y^{\binom{n}{2}-i} x^i \right]}{\partial x \partial y}$$

$$= -\sum_{i=0}^{\binom{n}{2}} I_n(i) \frac{\partial^2 \left[(y-x)^n y^{\binom{n}{2}-i} x^i \right]}{\partial x \partial y}$$

$$= (y-x)^n \sum_{i=0}^{\binom{n}{2}} I_n(i) y^{\binom{n}{2}-i} x^i$$

$$\times \left[\frac{2\binom{n}{2}}{(y-x)^2} + \frac{n(\binom{n}{2}-i)}{(y-x)y} - \frac{ni}{(y-x)x} - \frac{(\binom{n}{2}-i)i}{yx} \right].$$

□

Now using Lemma 4.6 and Corollary 4.9 we are ready to prove Theorem 4.3.

Proof. (*Proof of Theorem 4.3*) It is clear that the probability density function of $W_n = G(X_{n:n}) - G(X_{1:n})$ can be obtained from the joint probability density function of $G(X_{1:n})$ and $G(X_{n:n}) - G(X_{1:n})$ using the linear transformation. Denote $G(X_{1:n}) = Z_1$ and $G(X_{n:n}) = Z_2$. Using transformation $T_1 = Z_1$, $T_2 = Z_2 - Z_1$,we have $Z_1 = T_1$, $Z_2 = T_1 + T_2$. The Jacobian of this transformation equals to 1 and therefore

$$f_{T_1,T_2}(y_1, y_2) = f_{Z_1,Z_2(y_1,y_1+y_2)|J|^{-1}}$$

$$f_{T_2}(w) = \int_0^{1-w} f_{Z_1,Z_2}(y_1, y_1 + y_2) dy_1$$

$$f_{W_n}(w) = \int_0^{1-w} f_{G(X_{1:n}),G(X_{n:n})}(x, x + w) dx$$

$$f_{W_n}(w) = \int_0^{1-w} h(x, x+w)dx$$

$$= \sum_{i=0}^{\binom{n}{2}} I_n(i) \int_0^{1-w} (x+w)^{\binom{n}{2}-i} x^i \left[2\binom{n}{2} w^{n-2} + \frac{n\left(\binom{n}{2}-i\right)w^{n-1}}{(x+w)} - \frac{niw^{n-1}}{x} - \frac{\left(\binom{n}{2}-i\right)iw^n}{(x+w)x} \right] dx$$

$$= I_n(0) \left[\frac{2\binom{n}{2}}{\left(\binom{n}{2}+1\right)} w^{n-2}(1 - w^{\binom{n}{2}+1}) + nw^{n-1}\left(1 - w^{\binom{n}{2}}\right) \right]$$

$$+ \sum_{i=1}^{\binom{n}{2}} I_n(i) \left\{ 2(-i)^{-(i+1)} \binom{n}{2} w^{\binom{n+1}{2}-1} Beta\left[1 - \tfrac{1}{w}, i+1, \binom{n}{2} - i + 1\right] \right.$$

$$+ n(-1)^{-(i+1)} \left(\binom{n}{2} - i\right) w^{\binom{n+1}{2}-1} Beta\left[1 - \tfrac{1}{w}, i+1, \binom{n}{2} - i\right]$$

$$- n(1-w)^i w^{(n-2)} \ {}_2F_1\left[1, 1+\binom{n}{2}, i+1, 1-\tfrac{1}{w}\right]$$

$$\left. + i(-1)^{-(i+1)} \left(\binom{n}{2} - i\right) w^{\binom{n+1}{2}-1} Beta\left[1 - \tfrac{1}{w}, i, \binom{n}{2} - 1\right] \right\}$$

for $n \geq 1$, $0 \leq w \leq 1$.

Thus the theorem is proved. □

4.2 Some numerical results

4.2.1 Moments of the distribution of W

In Theorem 4.3 for special case of $F_i = G^i$, we have obtained the expression of $f_{W_n}(w)$, the pdf of the limiting distribution $P\{W_{1n} \leq x\} = P\{G(X_{n:n}) - G(X_{1:n}) \leq x\}$ which is given in (4.8). This pdf presents an independent interest and below we provide some numerical results and graphs concerning the numerical characteristics, as first, second and third moments, variance and skewness of the distribution.

n	$E(W_n)$	$E(W_n^2)$	$E(W_n^3)$	$Var(W_n)$	$Skewness$
2	0.333333	0.166667	0.100000	0.0555556	0.565685
3	0.466667	0.269048	0.175000	0.0512698	0.137187
4	0.529293	0.326647	0.222161	0.0464964	0.00484184
5	0.562734	0.360383	0.251653	0.0437131	−0.0377184
6	0.582385	0.381309	0.270722	0.0421366	−0.0492966
7	0.594799	0.395001	0.283555	0.0412154	−0.0502751
8	0.603094	0.404379	0.292525	0.0406568	−0.0480019
9	0.60889	0.411053	0.299006	0.0403066	−0.0451022
10	0.613089	0.415959	0.303826	0.0400806	−0.0424507

Table 2. Moments, variance and skewness of $f_{W_n}(w)$ for $n = 2, \ldots, 10$.

Below we provide the graphs of the pdf and cdf of W_n for different values of n.

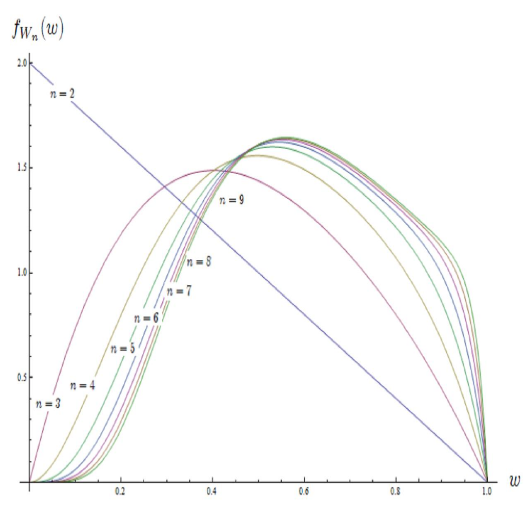


Figure 1. The graphs of f_{W_n} for $n = 2, \ldots, 9$.

Since $f_{W_2}(w) = 2 - 2w$, as can be seen from the Figure 1 for $n = 2$ the pdf of

W is linear.

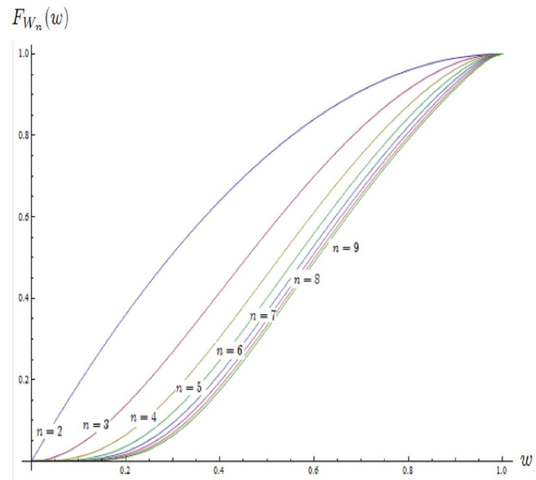


Figure 2. The graphs of F_{W_n} for $n = 2, \ldots, 9$.

Figure 3 and Figure 4 below provides the graphs of $E(W_n), E(W_n^2), E(W_n^3)$ and skewness with respect to n.

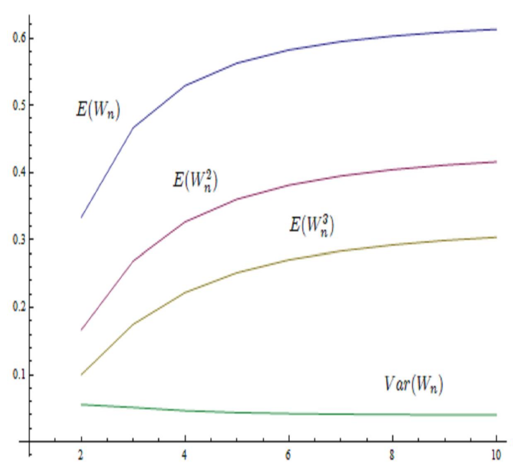


Figure 3. The graphs of moments and variance with respect to n.

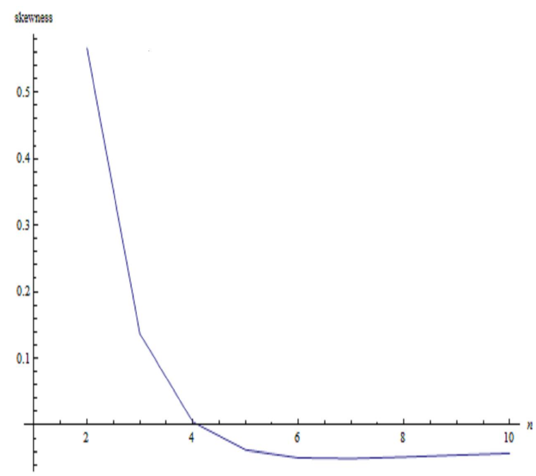


Figure 4. The graph of skewness with respect to n.

Additionally, from Figure 2-4 and Table 1 it can be seen that the moments $E(W_n), E(W_n^2), E(W_n^3)$ increase as n increases. The graph of $f_{W_n}(w)$ is right skewed for $n = 2, 3$ and left skewed for $n > 3$.

4.2.2 An application in insurance models

Suppose that the stationary series $\{Y_i,\ i = 1, 2, \ldots\}$ exhibits long-range dependence and has a marginal distribution function G with upper endpoint $y^G = sup\{y : G(y) < 1\}$. Interest is in upper extreme values of $\{Y_i\}$, which we take to be those that exceed a threshold u such that $1 - G(u)$ is small. In his study of emergency markets, [72] noted that Latin American markets have significantly fatter tails than industrial markets. So, depending on the chosen threshold value u, the industrial markets data coming from regions with stable economics rarely show extreme realizations. However, the finance data generally exhibit extreme observations. The key to deriving all statistical models for such extreme values is the asymptotic theory as $u \to y^G$ (see [23], Chapter 2). The most widely used approach is peaks over threshold method. A motivation of using INID random variables to model the extremes can be found in [32]. The authors derive and illustrate the advantage of a new model for the distribution of a peaks over a threshold obtained under a subasymptotic threshold $u < y^G$. The suggested exceedance modeling involves identifying independent clusters (with possible dependent extremes within the cluster) over u, selecting a cluster maxima X_i in the ith cluster and fitting these by independent Pareto random variables.

The exceedance models are used also in the modeling of insurance claims, in the case of existence of observations exceeding predefined thresholds in the portfolios. In general, under the assumption that the claim sizes are iid random variables, it is common to use the loss distributions which fit the observed data well, for example, the lognormal, Weibull, Pareto, Generalized Pareto. However, these loss distributions fail to fit the data in the case of the existence of long tails in exceedance models. Since insurance claims data generally includes large extreme observations and may not be identically distributed, modeling them with INID random variables is more realistic than the iid case. Focusing on the exceedances and using the proper distribution in models can provide a new and efficient approach. The exact distribution and the asymptotic properties of the number of observations falling near the maximum for iid case have been studied by [66], [53], [65], [43] and [19].

[41] describes a claim reinsurance treaties model in which $X_1, X_2, \ldots, X_k, \ldots$ are INID random variables that denote the claim sizes arising from a specific portfolio with continuous cumulative distribution functions $F_1, F_2, \ldots, F_k, \ldots$, respectively. Let $X_{1:k} \leq X_{2:k} \leq \cdots \leq X_{k:k}$ be the ordered values of claim sizes in the given portfolio corresponding to the random sample $X_1, X_2, \ldots, X_{k.}$ Let $\{N_1(t), t \geq 0\}$ be a stochastic process which counts the number of claims that occur in $[0, t], t \geq 0$ and we observe $X_1, X_2, \ldots, X_{N_1(t)}$ claims up to time $t > 0$. We assume that $N_1(t)$ is independent of $X_i's$. Let $Y_1, Y_2, \ldots, Y_m, \ldots$ be the sequence of independent and identically distributed random variables with the common continuous distribution function G and let us assume that given $N_2(t) = m$ the random sample $(Y_1, Y_2, \ldots, Y_{N_2(t)})$ is the unprocessed claim sizes in another specific portfolio, where $\{N_2(t), t > 0\}$ is another counting process independent of $Y's$. The following statistic $S_{N_2(t)}$ counts the number of claims from the second portfolio which falls between the lower and the higher claim

sizes in the first portfolio.

Denote

$$\xi_i = \begin{cases} 1, & Y_i \in (X_{1:N_1(t)}, X_{N_1(t):N_1(t)}) \\ 0, & \text{otherwise} \end{cases}$$

and

$$S_{N_2(t)} = \sum_{i=1}^{N_2(t)} \xi_i.$$

Therefore, conditioning on $N_1(t)$ and $N_2(t)$, a model for the second claim sizes can be represented by the probability density function obtained in [41]. The results can be useful in evaluating the asymptotic distributional properties of the number of claims whose size falls between the lower and higher thresholds of the first portfolio. For example, the data can be assumed to be INID with very special case of F^α scheme as in the paper. This assumption is supported with the consideration in practical applications saying that the distribution function of the first claim is greater than the distribution function of the second claim and so on. Mathematically, this can be written as $P\{X_1 \leq t\} \geq P\{X_2 \leq t\} \geq \cdots \geq P\{X_n \leq t\}, \ldots$ therefore the F^α scheme can be accomplished for this case.

Chapter 5

Conclusions

Exceedance models have been of great interest to researchers. Various aspects of exceedance statistics have been studied by the researches from different disciplines of science. Especially in quality control models, finance, insurance, modelling flood risk, etc. This study contributes to the growing body of research that addresses the impact of exceedances in models mentioned above.

In this dissertation, we explore the distributions of INID order statistics and the recurrence relations between them regarding the symmetric functions. Afterwards we deal with some difficulties of obtaining the limiting distribution of exceedances when they are from INID random variables. Based on the work of [3] the asymptotic distribution of exceedance statistics from INID random variables is derived considering F^{α}-scheme. Consideration of F^{α}-scheme in calculation of such problems contributes to the understanding of a INID random variables concept and also makes the calculations easier.

Moreover, the function we introduce in this study involves permutations with inversions and Gaussian hypergeometric function. So we give some detailed information based on permutations with inversions. The statistical characteristics of this function and graphical representations of them are derived and included in the study.

Appendix A

Mathematica Codes

A.1 Permutations with inversions

The following Mathematica code which provides the coefficients list of the n permutations with the k inversions is retrieved from the website (http://oeis.org/search?q=1%2C4%2C9%2C15%2C20%2C22&sort=&language=&go=Search) in the On-Line Encyclopedia of Integer Sequences (OEIS)

```
In[1]:= f[n_]:= CoefficientList[Expand@Product[Sum[x^i,
{i,0,j}],{j,n}],x]; Flatten[Array[f,9,0]]

nn=9;T[1,1]=1;T[n_,1]=0;

T[n_,k_]:=T[n,k]=Sum[T[n-i,k-1],{i,1,k-1}]

MatrixForm[Table[T[n,k],{n,nn},{k,nn}]] ;
```

So the coefficients are listed below:

```
In[2]:= A[1]={1};

In[3]:= A[2]={1,1};
```

```
In[4]:= A[3]={1,2,2,1};
```

```
In[5]:= A[4]={1,3,5,6,5,3,1};
```

```
In[6]:= A[5]={1,4,9,15,20,22,20,15,9,4,1};
```

```
In[7]:= A[6]={1,5,14,29,49,71,90,101,101,90,71,49,29,
14,5,1};
```

```
In[8]:= A[7]={1,6,20,49,98,169,259,359,455,531,573,
573,531,455,359,259,169,98,49,20,6,1};
```

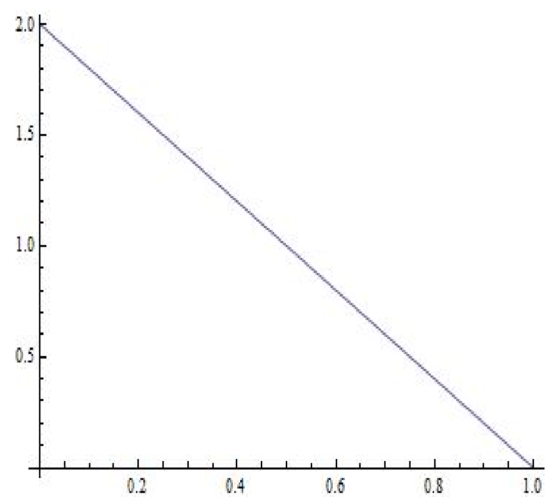
```
In[9]:= A[8]={1,7,27,76,174,343,602,961,1415,1940,2493,
3017,3450,3736,3836,3736,3450,3017,2493,1940,1415,961,
602,343,174,76,27,7,1};
```

```
In[10]:= A[9]={1,8,35,111,285,628,1230,2191,3606,5545,
8031,11021,14395,17957,21450,24584,27073,28675,29228,
28675,27073,24584,21450,17957,14395,11021,8031,5545,
3606,2191,1230,628,285,111,35,8,1};
```

A.2 Code for the probability density function $f_W(n)$

```
In[11]:= e[n_, i_]:=Integrate[w^n(n (n - 1)/w^2 x^i(w + x)
^(n(n-1)/2 - i)+n(n(n-1)/2-i)/wx^i(w + x)^(n(n-1)/2-i-1)
-ni/wx^(i-1)(w+x)^(n(n-1)/2-i)-i(n(n-1)/2-i)x^(i-1)(w+x)
^(n(n-1)/2-i-1)),{x,0, 1 - w},Assumptions->n(n-1)/2 >= i
 > 0 &&n>1 && 0<w<1]
```

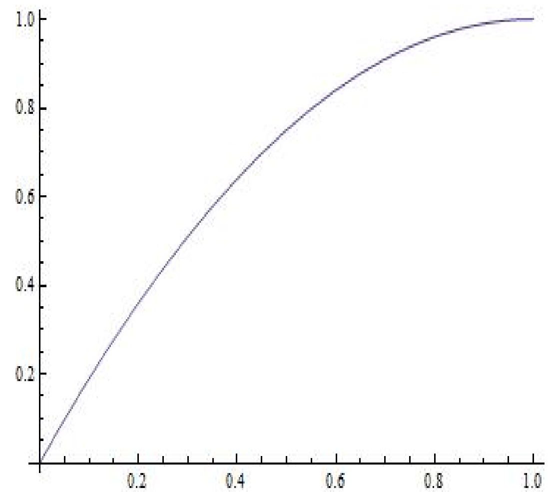
```
In[12]:= k[n_]:=w^(n)(n(n-1)/((n(n-1)/2+1)w^2)(1-w
^(n(n-1)/2+1))+n/w(1-w^(n(n-1)/2)))

In[13]:= r[n_]:=Expand[k[n]+Sum[A[n][[i+1]]e[n,i],
{i,1,Binomial[n,2]}]]

In[14]:= r[1]

Out[14]= 0

In[15]:= r[2]

Out[15]= 2-2w

In[16]:= Plot[r[2],{w,0,1}]
Out[16]:=
```



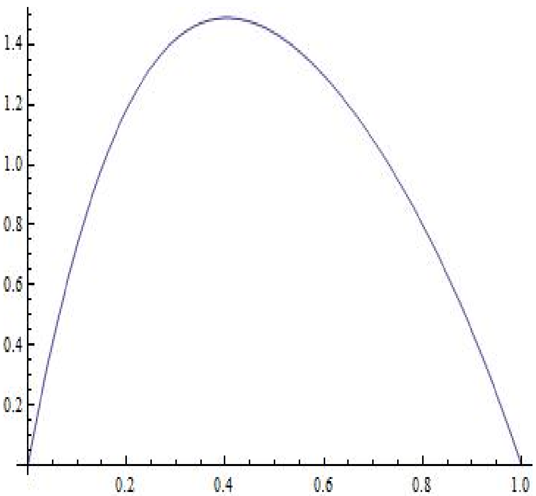

```
In[17]:= Integrate[r2[2],{w,0,s}]

Out[17]:= 2s-s^2
```

```
In[18]:=  Plot[Integrate[r2[2],{w,0,s}],{s,0,1}]
```

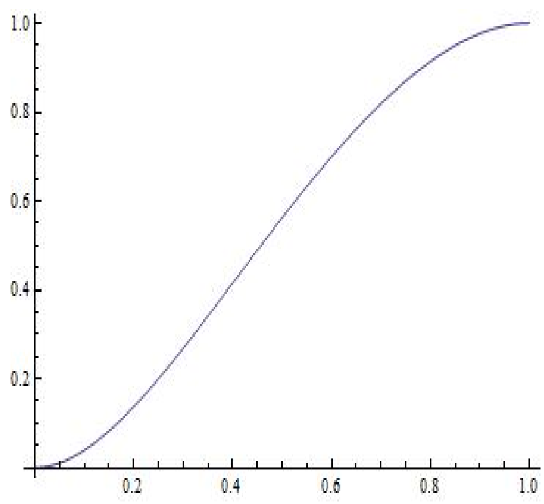
Out[18]=



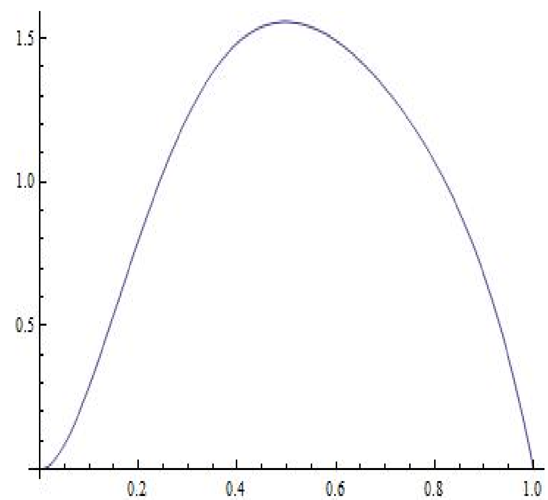

```
In[19]:=  r3[3]
Out[19]=  9w-18w^2+14w^3-5w^4

In[20]:=  Plot[r3[3],{w,0,1}]
Out[20]=
```




```
In[21]:= Integrate[r3[3],{w,0,s}]

Out[21]= (9s^2)/2-6s^3+(7s^4)/2-s^5

In[22]:= Plot[Integrate[(r3[3],{w,0,s}],{w,0,1}]
Out[22]=
```

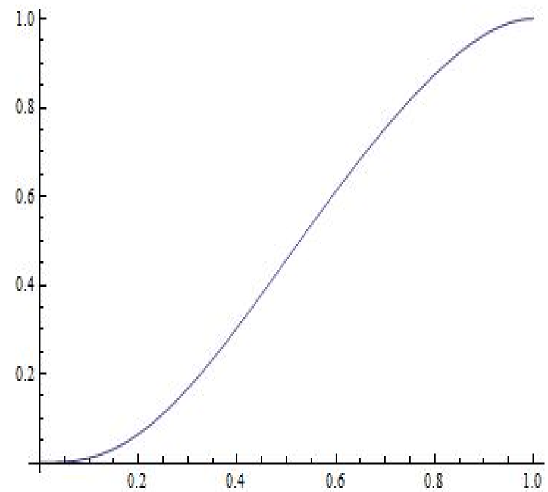


```
In[23]= r4[4]

Out[23]= (288w^2)/7-144w^3+228w^4-210w^5+(364w^6)/
3-44w^7+9w^8-(31w^9)/21

In[24]:= Plot[r4[4],{w,0,1}]

Out[24]=
```



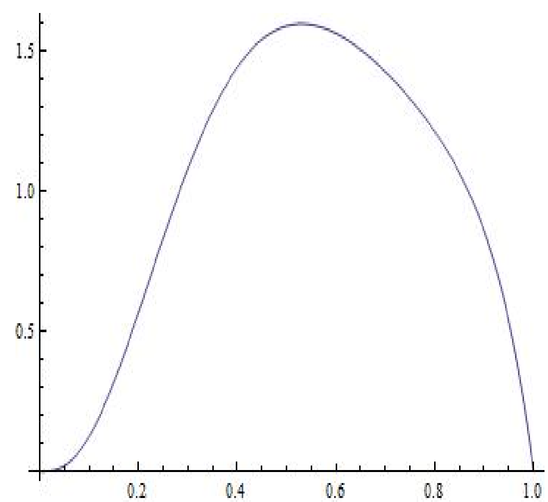

```
In[25]:= Integrate[r4[4],{w,0,s}]

Out[25]= (96s^3)/7-36s^4+(228s^5)/5-35s^6+(52s^7)/3
-(11s^8)/2+s^9-(31s^10)/210

In[26]:= Plot[Integrate[r4[4],{w,0,s}],{s,0,1}]
```

Out[26]=

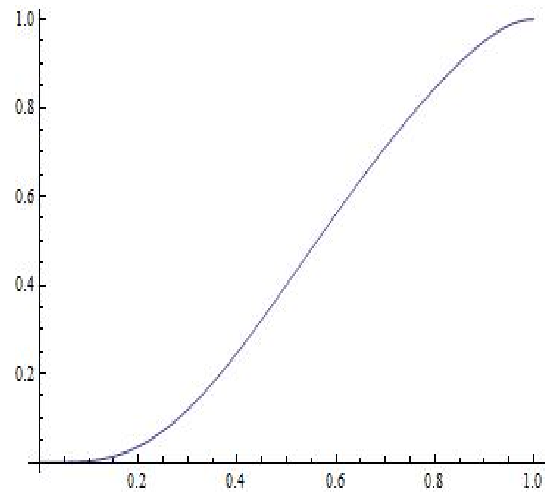


In[27]:= Plot[r5 [5],{ w,0,1}]

Out[27]=

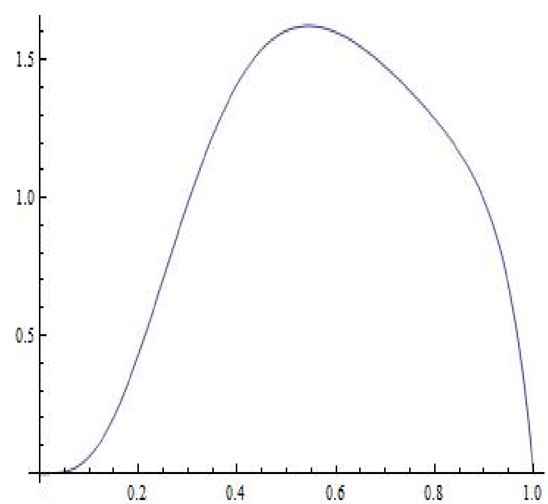


In[28]:= Plot[Integrate[r5 [5],{ w,0,s }],{ s ,0,1}]
Out[28]=



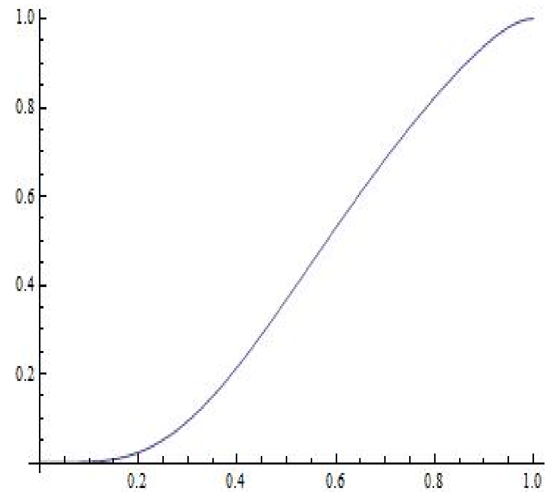
In[29]:= Plot[r6 [6],{ w,0,1}]

Out[29]=



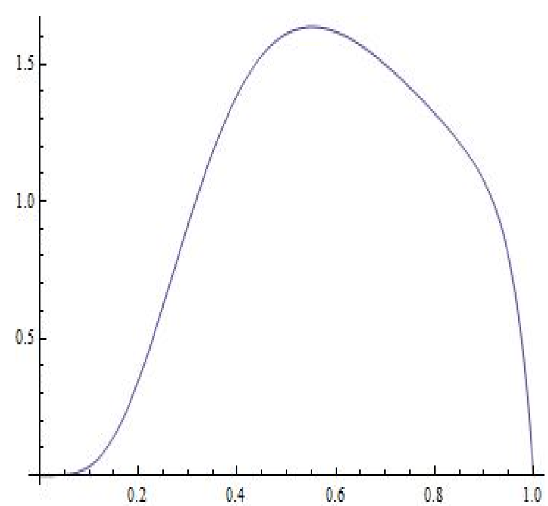
In[30]:= Plot[Integrate[r6 [6],{ w,0,s }],{ s ,0,1}]

Out[30]=



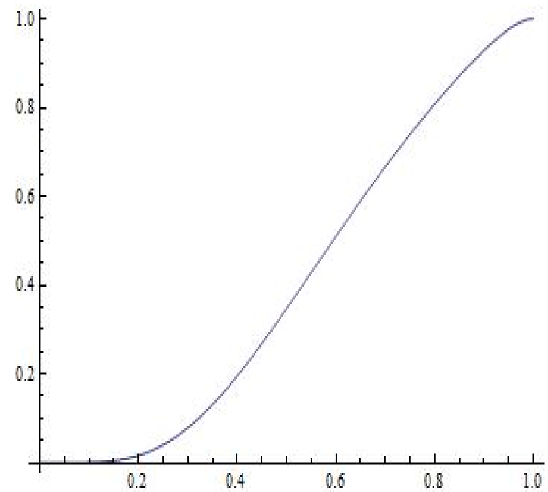

```
In[31]:= Plot[r7 [7],{ w,0,1}]
```

Out[31]=



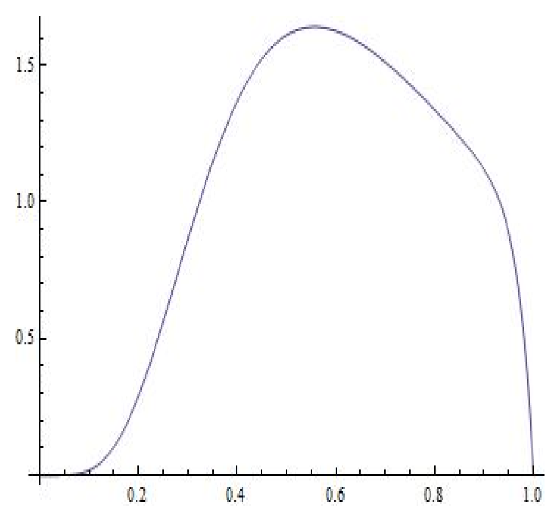

```
In[32]:= Plot[Integrate[r7 [7],{ w,0,s }],{ s ,0,1}]
```

Out[32]=



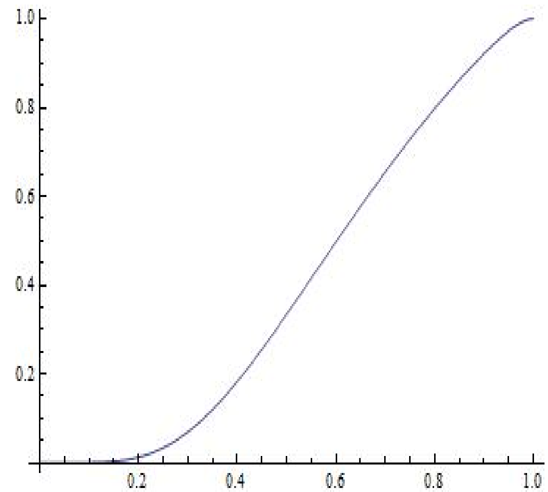

```
In[33]:= Plot[r8 [8],{ w,0,1}]
```

Out[33]=



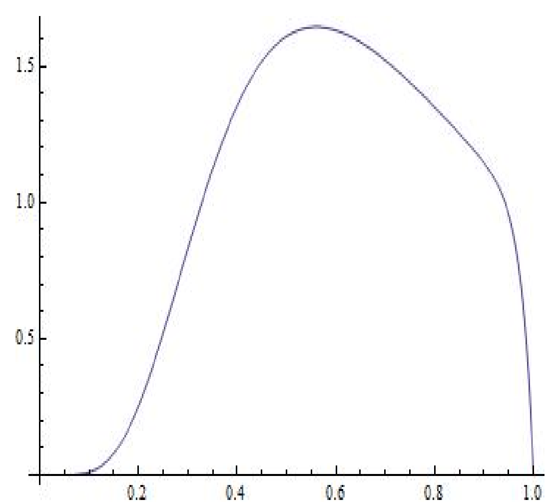

```
In[34]:= Plot[Integrate[r8 [8],{ w,0,s }],{ s ,0,1}]
```

Out[34]=

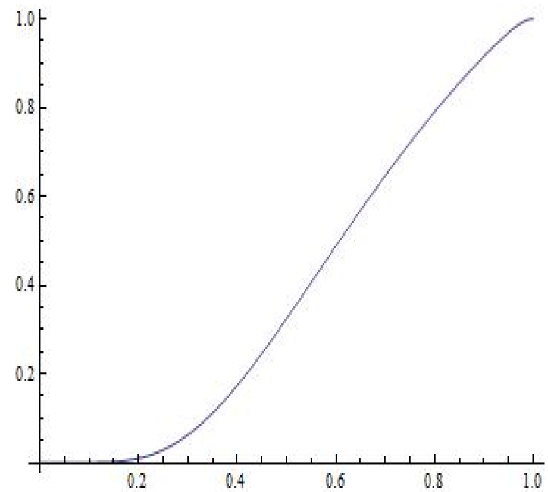


In[35]:= Plot[r9 [9],{ w,0,1}]

Out[35]=



In[36]:= Plot[Integrate[r9 [9],{ w,0,s }],{ s ,0,1}]
Out[36]=



A.3 Expected value of $f_W(w)$

```
In [37]:=  ex [ n_ , i_ ]:= Integrate [w  Integrate [(wˆn
(n(n−1)/wˆ2  xˆi(w+x)ˆ(n(n−1)/2−i)+n(n(n−1)/2−i)
/wxˆi(w+x)ˆ(n(n−1)/2−i −1)−ni/w  xˆ(i −1)(w+x)
ˆ(n(n−1)/2−i)−i(n(n−1)/2−i)xˆ(i −1)(w+x)ˆ(n(n−1)
/2−i −1))),{x,0,1−w},Assumptions−> n(n−1)/2 >=i>0
 && n>1 && 0<w<1],{w,0 ,1}]

In [38]:=  kex [ n_ ]:= Integrate [w(wˆ(n)(n(n−1)/
((n(n−1)/2+1)wˆ2)(1−wˆ(n(n−1)/2+1))+n/w(1−w
ˆ(n(n−1)/2)))),{w,0 ,1}]

In [39]:=  ex [ n_ ]:=N[Expand [ kex [n]+Sum[A[n]
[[ i+1]]ex[n, i ] ,
{i ,1 , Binomial [n ,2] } ] ] ]
```

```
In[40]:= ex[1]
Out[40]= 0
```

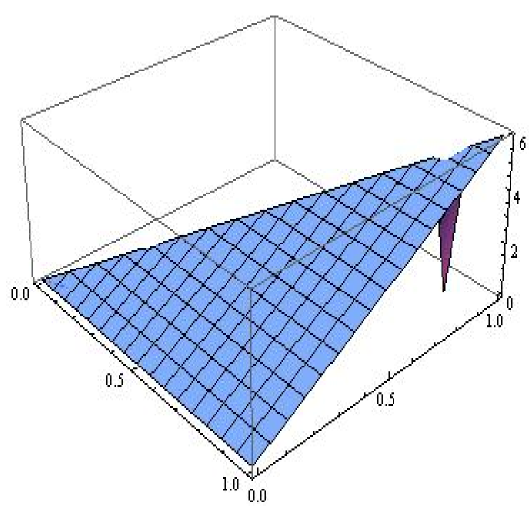
```
In[41]:= ex[2]
Out[41]= 0.333333

In[42]:= ex[3]
Out[42]= 0.466667

In[43:= ex[4]
Out[43]= 0.529293

In[44]:= ex[5]
Out[44]= 0.562734

In[45]:= ex[6]
Out[45]= 0.582385

In[46]:= ex[7]
Out[46]= 0.594799

In[47]:= ex[8]
Out[47]= 0.603094

In[48]:= ex[9]
Out[48]= 0.60889
```

A.4 $h(x,y)$ and $H(x,y)$ and their plots

```
In[49]:= e1[n_,i_]:=(y-x)^n(n(n-1)/(y-x)^2x^i(y)
^(n(n-1)/2-i)+n(n(n-1)/2-i)/(y-x)x^i(y)^(n(n-1)
```

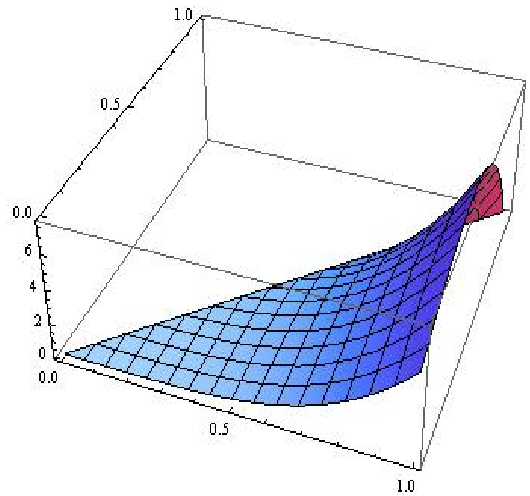
```
/2−i −1)−ni/(y−x)x^(i −1)(y)^(n(n−1)/2−i)−i(n(n−1)
/2−i)x^(i −1)(y)^(n(n−1)/2−i −1))

In[50]:= Integrate[x y e1[n,i],{y,0,1},{x,0,y},
Assumptions −> n(n−1)/2>=i> 0]

In[51]:= pdfh[n_]:=Simplify[Sum[A[n][[i+1]]e1[n,i],
{i,0,Binomial[n,2]}]]

In[52]:= intpdfh[n_]:=Simplify[Integrate[Sum[A[n]
[[i+1]]e1[n,i],{i,0,Binomial[n,2]}],{y,0,1},{x,0,y}]]

In[53]:= pdfh[2]

Out[53]= 2 (x + y)

In[54]:= Plot3D[pdfh[2],{y,0,1},{x,0,y}]
```

Out[54]=



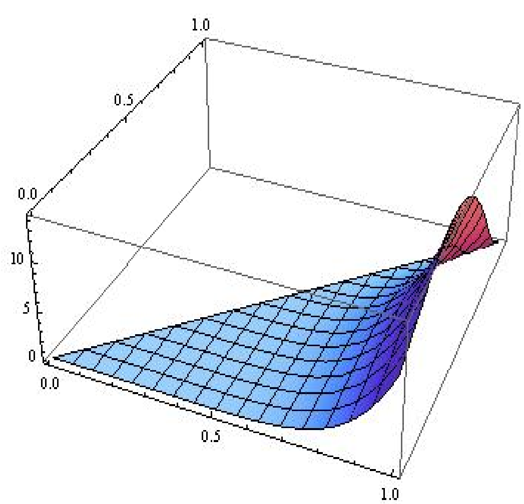
```
In[55]:= Plot3D[pdfh[3],{y,0,1},{x,0,y}]
```

Out[55]=




```
In[56]:= Plot3D[pdfh[4],{y,0,1},{x,0,y}]
```

Out[56]=



BIBLIOGRAPHY

[1] M. Ahsanullah, *Record statistics*, Nova Science Publishers, Inc., Commack, NY, 1995. MR1443904

[2] B. C. Arnold, N. Balakrishnan and H. N. Nagaraja, *A first course in order statistics*, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons Inc., New York, A Wiley-Interscience Publication, 1992. MR1178934

[3] I. G. Bairamov, *Some distribution free properties of statistics based on record values and characterizations of distributions through a record*, Journal of Applied Statistical Science **5** (1997), 17–25. MR1439625

[4] I. G. Bairamov, *Advances in Exceedance Statistics Based on Ordered Random Variables*, In Recent developments in ordered random variables, Nova Science Publishers, New York, 97–117, 2007. MR2412839

[5] I. G. Bairamov and S. N. Eryilmaz, *Distributional properties of statistics based on minimal spacing and record exceedance statistics*, Journal of Statistical Planning and Inference **90** (2000), 21–33. MR1791579

[6] I. G. Bairamov and S. N. Eryilmaz, *Applied Statistical Sciences: On properties of statistics connected with minimal spacing and record exceedances*, Ahsanullah, M., Kennyon, J. and Sarkar, S. K. (Editors). Nova Science Publishers, Inc., Huntington, New York, **V**, 245–254, 2001.

[7] I. G. Bairamov and S. N. Eryilmaz, *Characterization of symmetry and exceedance models in multivariate FGM distributions*, Journal of Applied Statistical Science **13** (2004), 87-99. MR2116788

[8] I. G. Bairamov and S. N. Eryilmaz, *Spaciings, exceedances and concomitants in progressive type II censoring scheme*, Journal of Statistical Planning and inference **136** (2006), 527–536.

[9] I. G. Bairamov and S. N. Eryilmaz, *Joint Behaviour of precedences and exceedances in random threshold models*, Australian and New Zealand Journal of Statistics **50** (2008), 209–216

[10] I. G. Bairamov and S. N. Eryilmaz, *Waiting times of exceedances in random threshold models*, Statist. and Probab. Letters **79** (2009), 676–683.

[11] I. G. Bairamov and O. L Gebizlioglu O.L. *On the distance sequence of distribution functions*, Journal of Applied Statistical Science. **8** (1998), 59–66.

[12] I. G. Bairamov and M. K. Khan, *On exceedances of record and order statistics*, Proceedings of the American Mathematical Society **135** (2007), 1935–1945. MR2286107

[13] I. G. Bairamov and S. Kotz, *On distributions of exceedances associated with order statistics and record values for arbitrary distributions*, Statistical Papers **42** (2001), 171–185. MR1836847

[14] I. G. Bairamov and H. Tanil, *On distributions of exceedances based on generalized order statistics*, Communications in StatisticsTheory and Methods **36** (2007), 1479–1491.

[15] I. G. Bairamov and H. Tanil, *Distributions of exceedances of generalized order statistics*, Statistics **42** (2008), 67–76.

[16] I. G. Bairamov and Y. I. Petunin, *The structure of invariant confidence intervals that contain the basic distributed mass*, (Russian) Teor. Veroyatnost. i Primenen **35** (1990), 15–26; translation in Theory Probab. Appl. **35** (1990), 15–26. MR1050050

[17] N. Balakrishnan, *Permanent, Order Statistics, Outliers and Robustness, Review*, Mathematics of Computation **1** (2007), 7–107. MR2310579

[18] N. Balakrishnan and H. K. T. Ng, *Precedence-type tests and applications*, Wiley Series in Probability and Statistics. Wiley-Interscience, John Wiley and Sons, Hoboken, NJ, 2006. MR2217899

[19] N. Balakrishnan, A. G. Pakes, A. Stepanov, *On the number and sum of near-record observations*, Advances in Applied Probability **37** (2005), 765–780. MR2156559

[20] N. Balakrishnan and C. R. Rao, *Order statistics: theory and methods*, Handbook of Statistics, 16. North-Holland Publishing Co., Amsterdam, 1998. MR1668739

[21] N. Balakrishnan and C. R. Rao, *Order statistics: applications*, Handbook of Statistics, 17. North-Holland Publishing Co., Amsterdam, 1998. MR1672283

[22] H. Bateman, *Higher Transcendental Functions*, MacGraw Hill book Company, INC., New York, USA, 1953. MR58756

[23] J. Beirlant, J. Goegebeur, J. Teugels and J. Segers *Statistics of Extremes: Theory and applications. With contributions from Daniel De Waal and Chris Ferro*, Wiley Series in Probability and Statistics. John Wiley and Sons, Ltd., Chichester, 2004. MR2108013

[24] J. P. M. Binet, *Mé moire sur la thé orie des axes conjugué s et des moments d'inertie des corps*, Journal de l'É cole Polytechnique **9** (1813), 280–302.

[25] J. Bourget, *Des Permutations*, Nouvelles Anna les De Mathematiques **10** (1871), 254–268.

[26] A. L. Cauchy, *Sur les fonctions qui ne peuvent obtenir que deux valeurs é gales et de signes contraires par suite des transpositions opé ré es entre les variables qu'elles renferment*, Journal de l'É cole Polytechnique **10** (1812), 29–112.

[27] L. Comtet, *Advanced Combinatorics,enlarged ed.*, D. Reidel Publishing Company, Dordrecht, 1974. MR0460128

[28] H. A. David, *Order statistics*, **2** , John Wiley and Sons Inc., New York, 1981. MR0597893

[29] H. A. David and H. N. Nagaraja, *Order statistics*, **3**, Wiley Series in Probability and Mathematical Statistics, Wiley-Interscience, John Wiley and Sons Inc., Hoboken, NJ, 2003. MR1994955

[30] P. Embrechts, C. Klppelberg, T. Mikosch, *Modelling extremal events. For insurance and finance. Applications of Mathematics*, **33**, Springer-Verlag, Berlin, 1997. MR1458613

[31] B. Epstein, *Tables for the distribution of the number of exceedances*, Annals of Mathematical Statistics **25** (1954), 762–768. MR65074

[32] E. Eastoe and J. Tawn, *Modelling the distribution of cluster maxima of exceedances of sub asymptotic thresholds*, Biometrika **99** (2012), 43–55.

[33] D. G. J. Farlie, *The performance of some correlation coefficients for a general Bivariate distribution*, Biometrika **47** (1960), 307–323.

[34] J. Galambos, *The asymptotic theory of extreme order statistics*, Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York-Chichester-Brisbane, 1978. MR489334

[35] B. V. Gnedenko, B.N. *The theory of Probability*, Translated from the Russian by G. Yankovskii, Mir Publishers, Moskow, 1978. MR0524397

[36] E. J. Gumbel, *Statistics of extremes*, Columbia University Press, New York 1958. MR96342

[37] E. J. Gumbel, *Bivariate exponential distributions*, J. Am. Stat. Assoc. **55** (1960), 698–707. MR81575

[38] E. J. Gumbel and H von Schelling, *The distribution of the number of exceedances*, Ann. Math. Statistics **21** (1950), 247-262. MR35415

[39] S. Gurler and I. Bairamov, *Parallel and k-out-of-n: G systems with nonidentical components and their mean residual life functions*, Applied Mathematical Modelling **33** (2009), 1116–1125. MR2468510

[40] H. L. Harter, *The chronological Annotated Bibliography of Order Statistics, (Vol. 1-8)*, American Sciences Press, Columbus, Ohio, 1978–1992.

[41] E. Hashorva, *On the asymptotic distribution of certain bivariate reinsurance*, Insurance: Mathematics and Economics **40** (2007), 200–208.

[42] M. Henryk, *Permanents, with a foreword by Marvin Marcus, Encyclopedia of Mathematics and its Applications*, **6**, Addison-Wesley Publishing Co., Reading, 1978.

[43] Z. Hu, C. Su, *Limit theorems for the number and sum of near-maxima for medium tails*, Statistics and Probability Letters **63** (2003), 229–237. MR1986321

[44] N. L. Johnson and S. Kotz *Some generalizations of Bernoulli and Plya-Eggenberger contagion models*, Statist. Hefte **32** (1991), 1–17. MR1123680

[45] N. L. Johnson and S. Kotz *On some generalized farlie-gumbel-morgenstern distributions*, Communications in Statistics: Theory and Methods **4** (1975), 415–427.

[46] N. L. Johnson and S. Kotz *On some generalized farlie-gumbel-morgenstern distributions-II regression, correlation and further generalizations*, Communications in Statistics: Theory and Methods **6** (1977), 485–496.

[47] N. L. Johnson and S. Kotz *Further comments on Matveychuk and Petunin's generalized Bernoulli model, and nonparametric tests of homogeneity*, J. Statist. Plann. Inference **41** (1994), 61–72. MR1292147

[48] N. L. Johnson, S. Kotz, A. W. Kemp, *Univariate discrete distributions. Second edition*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1992. MR1224449

[49] W. Katzenbeisser, *The distribution of two-sample location exceedance test statistics under Lehmann alternatives*, Statist. Hefte **26** (1985), 131–138. MR861796

[50] W. Katzenbeisser, *The exact power of two-sample location tests based on exceedance statistics against shift alternatives*, Statistics **20** (1989), 47–54. MR995233

[51] D. E. Knuth, *The Art of Computer Programming*, **3**, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont. Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing, 1973. MR445948

[52] M. R. Leadbetter, G. Lindgren and H. Rootzn, *Extremes and related properties of random sequences and processes*, Springer Series in Statistics. Springer-Verlag, New York-Berlin, 1983. MR691492

[53] Y. Li, *A note on the number of records near the maximum*, Statistics and Probability Letters **43** (1999), 153–158. MR1693281

[54] I. G. Macdonald, *Symmetric functions and Hall polynomials. Second edition. With contributions by A. Zelevinsky*, Oxford Mathematical Monographs, Oxford Science Publications, The Clarendon Press, Oxford University Press, New York, 1995. MR1354144

[55] B. H. Margolius, *Permutations with inversions*, Journal of Integer Sequences **4(2)** (2001), Article 01.2.4 (electronic). MR1873402

[56] S. A. Matveichuk and Y. I. Petunin, *Generalization of Bernoulli schemes that arise in order statistics*, (Russian) Ukrain. Mat. Zh. **42** (1990), 518–528; translation in Ukrainian Math. J. **42** (1990), 459–466 MR1051875

[57] D. Morgenstern, *Einfache Beispiele zweidimensionaler Verteilungen*, Mitteilingsblatt fur Mathematische Statistik **8** (1956), 234–235 MR81575

[58] T. Muir, *A treatise on the theory of determinants with graduate sets of exercises for use in colleges and schools*, MacMillian and Co., London, UK 1882.

[59] T. Muir, *On a simple term of a determinant*, Proceedings of the Royal Society of Edinburgh **21** (1898), 441–477.

[60] W. Maurer and B. H. Margolin, *The Multivariate inclusion-exclusion formula and order statistics from dependent variates*, Annals of Statistics **4** (1976), 1190–1199. MR0426287

[61] H. N. Nagaraja, *Record Values and Related Statistics a Review*, Commun. Statist. Theory Meth. **17** (1988), 2223–2238.

[62] V. B. Nevzorov, *The record and interrecord times for sequences of nonidentically distributed random variables*, Zapiski Nauchnykh Seminarov LOMI **142** (1985), 109–118. MR0788192

[63] V. B. Nevzorov, *Two characterizations using records. In Stability problems for stochastic models*, Lecture Notes in Mathematics, Springer, Berlin **1233** (1987), 79–85. MR0886283

[64] V. B. Nevzorov, *Records*, Theo. Prob. Apl. **32** (1988), 201–228.

[65] A. Pakes, *The number and sum of near-maxima for thin-tailed populations*, Advances in Applied Probability **32** (2000), 1100–1116. MR1808916

[66] A. Pakes, F. W. Steutel, *On the number of records near the maximum*, Australian Journal of Statistics **39** (1997), 179–193. MR1481260

[67] D. Pfeifer, *Extremal processes, secretary problems, and the l/e law*, Journal of Applied Probability **26** (1989), 722–733. MR1025389

[68] D. Pfeifer, *Some remarks on Nevzorov's record model*, Advances in Applied Probability **23** (1991), 823–833. MR1133730

[69] R. D. Reiss, *Approximate distributions of order statistics. With applications to nonparametric statistics*, Springer Series in Statistics. Springer-Verlag, New York, 1989. MR988164

[70] S. I. Resnick, *Extreme values, regular variation, and point processes*, Applied Probability. A Series of the Applied Probability Trust, **4**, Springer-Verlag, New York, 1987. MR900810

[71] H. Robbins, *On distribution-free tolerance limits in random sampling*, Ann. Math. Statistics **15** (1944), 214–216. MR10365

[72] R. Samuel, *Extreme observations and diversification in emerging markets*, Journal of International Money and Finance **20** (2001), 971–986.

[73] A. E. Sarhan and B. G. Greenberg, *Contributions to order statistics*, John Wiley and Sons, Inc., New York-London, 1962. MR169336

[74] K. Sarkadi, *On the distribution of the number of exceedances*, Ann. Math. Stat. **28** (1957), 1021–1023.

[75] M. M. Siddiqui, *Order statistics for a sample and extended sample*, Non-parametric Techniques in Statistical Inference, Cambridge University Press, Cambridge, (1970), 417–423.

[76] R. P. Stanley, *Enumerative combinatorics. Vol. 2. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin*, Cambridge Studies in Advanced Mathematics, 62. Cambridge University Press, Cambridge, 1999. MR1676282

[77] I. Todhunter, *A History of the Mathematical Theory of Probability*, Macmillan, Cambridge and London, 1865, Reprinted Z. by Chelsea, New York. Z., 1965.

[78] R. J. Vaughan and W. N. Venables, *Permanent expressions for order statistics densities*, Journal of the Royal Statistical Society: Series B **34** (1972), 308–310. MR0326919

[79] J. Wesolowski and M. Ahsanullah, *Distributional Properties of Exceedance Statistics*, Annals of the Institute of Statistical Mathematics, **50(3)** (1998), 543–565. MR1664524

[80] S. S. Wilks, *Determination of sample sizes for setting tolerance limits*, Ann. Math. Statistics **12** (1941), 91–96. MR4451

[81] S. S. Wilks, *Statistical prediction with special reference to the problem of tolerance limits*, Ann. Math. Statistics **13** (1942), 400-409. MR7592

[82] S. S. Wilks, *Recurrence of extreme observations*, J. Australian Math. Soc. **1** (1959), 106–112. MR7592

VITAE

Göknur Giner was born in Çorlu, Turkey, on August 1, 1981. She completed her high school education in İzmir. After she earned her bachelor's degree in statistics, she received her master degree of Operations Research from Marmara University, İstanbul, in 2005. She worked as a research assistant at the department of Mathematics at İzmir University of Economics from 2005 to 2008. Whilst there she received her master of science degree in theoretical statistics from Ege University, İzmir, in 2007. She also worked as a research assistant at the department of Statistics at Yasar University between 2008 and 2010 and at the department of Mathematics and Computing at Beykent University from 2010 to 2011. Recently, she lives and works within the statistics and mathematics community in Melbourne, Australia.