ATLANTIS PRESS

Atlantis Studies in Probability and Statistics
Series Editor: C.P. Tsokos

Muhammad Qaiser Shahbaz
Mohammad Ahsanullah
Saman Hanif Shahbaz
Bander M. Al-Zahrani

Ordered Random Variables: Theory and Applications

Atlantis Studies in Probability and Statistics

Volume 9

Series editor

Chris P. Tsokos, Tampa, USA

Aims and scope of the series

The series 'Atlantis Studies in Probability and Statistics' publishes studies of high quality throughout the areas of probability and statistics that have the potential to make a significant impact on the advancement in these fields. Emphasis is given to broad interdisciplinary areas at the following three levels:

(I) Advanced undergraduate textbooks, i.e., aimed at the 3rd and 4th years of undergraduate study, in probability, statistics, biostatistics, business statistics, engineering statistics, operations research, etc.;
(II) Graduate-level books, and research monographs in the above areas, plus Bayesian, nonparametric, survival analysis, reliability analysis, etc.;
(III) Full Conference Proceedings, as well as selected topics from Conference Proceedings, covering frontier areas of the field, together with invited monographs in special areas.

All proposals submitted in this series will be reviewed by the Editor-in-Chief, in consultation with Editorial Board members and other expert reviewers.

For more information on this series and our other book series, please visit our website at: www.atlantis-press.com/Publications/books

AMSTERDAM—PARIS—BEIJING
ATLANTIS PRESS
Atlantis Press
29, avenue Laumière
75019 Paris, France

More information about this series at http://www.atlantis-press.com

Muhammad Qaiser Shahbaz
Mohammad Ahsanullah
Saman Hanif Shahbaz
Bander M. Al-Zahrani

Ordered Random Variables: Theory and Applications

ATLANTIS
PRESS

Muhammad Qaiser Shahbaz
Department of Statistics
King Abdulaziz University
Jeddah
Saudi Arabia

Saman Hanif Shahbaz
Department of Statistics
King Abdulaziz University
Jeddah
Saudi Arabia

Mohammad Ahsanullah
Department of Management Sciences
Rider University
Lawrenceville, NJ
USA

Bander M. Al-Zahrani
Department of Statistics
King Abdulaziz University
Jeddah
Saudi Arabia

ISSN 1879-6893 ISSN 1879-6907 (electronic)
Atlantis Studies in Probability and Statistics
ISBN 978-94-6239-224-3 ISBN 978-94-6239-225-0 (eBook)
DOI 10.2991/978-94-6239-225-0

Library of Congress Control Number: 2016951690

Printed on acid-free paper

To my Father, Mother and my Wife Saman

Muhammad Qaiser Shahbaz

To my Father, Mother and my Wife, Masuda

Mohammad Ahsanullah

To my Father, Mother and my Husband Shahbaz

Saman Hanif Shahbaz

To my parents

Bander Al-Zahrani

Preface

Ordered random variables have attracted several researchers due to their applicability in many areas, like extreme values. These variables occur as a natural choice when dealing with extremes like floods, earthquakes, etc. The use of ordered random variables also appears as a natural choice when dealing with records. In this book we have discussed various models of ordered random variables with both theoretical and application points of view. The introductory chapter of the book provides a brief overview of various models which are available to model the ordered data.

In Chap. 2 we have discussed, in detail, the oldest model of ordered data, namely order statistics. We have given the distribution theory of order statistics when sample is available from some distribution function F(x). Some popular results regarding the properties of order statistics have been discussed in this chapter. This chapter also provides a brief about reversed order statistics which is a mirror image of order statistics. We have also discussed recurrence relations for moments of order statistics for various distributions in this chapter.

Chapter 3 of the book is dedicated to another important model of ordered variables, known as record values introduced by Chandler (1952). Record values naturally appear when dealing with records. This chapter discusses in detail the model when we are dealing with larger records and is known as upper record values. The chapter contains distribution theory for this model alongside some other important results. The chapter also presents recurrence relations for moments of record values for some popular probability distributions.

Kamps (1995) introduced a unified model for ordered variables, known as generalized order statistics (GOS). This model contains several models of ordered data as a special case. In Chap. 4, we have discussed, in detail, this unified model of ordered data. This chapter provides a brief about distribution theory of GOS and its special cases. The chapter also contains some important properties of the model, like Markov chains property and recurrence relations for moments of GOS for some selected distributions.

In Chap. 5 the model of reversed order random variables known as dual generalized order statistics (DGOS) is discussed. The model was introduced by Burkschat et al. (2003) as a unified model to study the properties of variables arranged in decreasing order. The model contains reversed order statistics and lower record values as a special case. We have given some important distributional properties for the model in Chap. 5. We have also discussed recurrence relations for moments of DGOS when sample is available from some distribution F(x). The chapter also provides relationship between GOS and DGOS.

Ordered random variables have found tremendous applications in many areas such as estimation and concomitants. Chapter 6 of the book presents some popular uses of ordered random variables. The chapter presents use of ordered random variables in maximum likelihood and Bayesian estimation.

Chapters 7 and 8 of the book present some popular results about probability distributions which are based on ordered random variables. In Chap. 7 we have discussed some important results regarding the characterization of probability distributions based on ordered random variables. We have discussed characterizations of probability distributions based on order statistics, record values, and generalized order statistics. Chapter 8 contains some important results which connect ordered random variables with extreme value distribution. We have discussed the domains of attractions for several random variables for various types of extreme values distributions.

Finally, we would like to thank our colleagues and friends for their support and encouragement during compilations of this book. We would like to thank Prof. Chris Tsokos for valuable suggestions which help in improving the quality of the book. Authors 1 and 3 would like to thank Prof. Muhammad Hanif and Prof. Valary Nevzorov for healthy comments during compilation of this book. Authors 1, 3, and 4 would also like to thank Statistics Department, King Abdulaziz University, for providing excellent support during compilation of the book. Author 2 would like to thank Rider University for their excellent facilities which helped in completing the book.

Jeddah, Saudi Arabia — Muhammad Qaiser Shahbaz
Lawrenceville, USA — Mohammad Ahsanullah
Jeddah, Saudi Arabia — Saman Hanif Shahbaz
Jeddah, Saudi Arabia — Bander M. Al-Zahrani

Contents

Chapter 1
Introduction

1.1 Introduction

Ordered Random Variables arise in several areas of life. We can see the application of ordered random variables in our daily life for example we might be interested in arranging prices of commodities or we may be interested in arranging list of students with respect to their CGPA in final examination. Another use of ordered random variables can be seen in games where; for example; record time in completing a 100 m race is recorded. Several such examples can be listed where ordered random variables are playing their role. The ordered random variables has recently attracted attention of statisticians although their use in statistics is as old as the subject. Some simple statistical measures which are based upon the concept of ordered random variables are the range, the median, the percentiles etc. The ordered random variables are based upon different models depending upon how the ordering is being done. In the following we will briefly discuss some popular models of ordered random variables which will be studied in more details in the coming chapters.

1.2 Models of Ordered Random Variables

Some popular models of ordered random variables are discussed in the following.

1.2.1 Order Statistics

Order Statistics is perhaps the oldest model for ordered random variables. Order Statistics naturally arise in life whenever observations in a sample are arranged in increasing order of magnitude. The order statistics are formally defined as under.

M.Q. Shahbaz et al., *Ordered Random Variables: Theory and Applications*,
Atlantis Studies in Probability and Statistics 9, DOI 10.2991/978-94-6239-225-0_1

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from some distribution function $F(x)$. The observations arranged in increasing order of magnitude $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ are called order statistics for a sample of size n. The joint distribution of all order statistics is given in David and Nagaraja (2003) as

$$f_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i). \tag{1.1}$$

Order statistics are very useful in studying distribution of maximum, minimum, median etc. for specific probability distributions. We will study order statistics in much detail in Chap. 2. The rth order statistics $X_{r:n}$ can be viewed as life length of $(n - r + 1)$–out–of–n system.

1.2.2 Order Statistics with Non-integral Sample Size

The study of order statistics is based upon size of the available sample and conventionally that sample size is a positive integer. The model of order statistics is easily extended to the case of fractional sample size to give rise to fractional order statistics as defined by Stigler (1977). The conventional order statistics appear as a special case of fractional order statistics. The distribution function of rth fractional Order Statistics based upon the parent distribution $F(x)$ is given as

$$F_{r:\alpha}(x) = \frac{1}{B(\alpha, r)} \int_0^{F(x)} t^{r-1} (1 - t)^{\alpha - r} \, dt. \tag{1.2}$$

If $\alpha = n$ (integer) then we have simple order statistics. The fractional order statistics do not have significant practical applications but they do provide basis for a general class of distributions introduced by Eugene, Lee and Famoye (2002).

1.2.3 Sequential Order Statistics

If we consider simple order statistics as life length of components then we can interpret them as random variables where the probability distribution of components remains same irrespective of the failures. In certain situations the probability distribution of the components changes after each failure and hence such components can not be modeled by using simple order statistics. Sequential Order Statistics provide a method to model the components with different underlying probability distributions after each failure. The Sequential Order Statistics are defined as follow.

Let $\left\{Y_j^{(i)}; i = 1, 2, \ldots, n; j = 1, 2, \ldots, n - i + 1\right\}$ be independent random variables so that $\left\{Y_j^{(i)}; j = 1, 2, \ldots, n - i + 1\right\}$ is distributed as F_i and $F_1, \ldots, F_n$ are strictly increasing. Moreover let $X_j^{(1)} = Y_j^{(1)}$, $X_*^{(1)} = \min\left\{X_1^{(1)}, \ldots, X_n^{(1)}\right\}$ and for $2 \le i \le n$ define

$$X_j^{(i)} = F_i^{-1}\left[F_i\left(Y_j^{(i)}\right)\left\{1 - F\left(X_*^{(i-1)}\right)\right\} + F_i\left(X_*^{(i-1)}\right)\right];$$
$$X_*^{(i)} = \min\left\{X_j^{(i)}, 1 \le j \le n - i + 1\right\}$$

then the random variables $X_*^{(1)}, \ldots, X_*^{(n)}$ are called Sequential Order Statistics. The joint density of first r Sequential Order Statistics; $X_*^{(1)}, \ldots, X_*^{(r)}$; is given by Kamps (1995b) as

$$f_{1,2,\ldots,r:n}(x_1, x_2, \ldots, x_r) = \frac{n!}{(n-r)!}\prod_{i=1}^{r}\left\{\frac{1 - F_i(x_i)}{1 - F_i(x_{i-1})}\right\}^{n-i} \times \frac{f_i(x_i)}{1 - F_i(x_{i-1})}. \tag{1.3}$$

The Sequential Order Statistics reduces to simple order statistics if all $F_i(x)$ are same.

1.2.4 Record Values

The Record values has emerged as an important model for ordered random variables. The record values appear naturally in real life where one is interested in successive extreme values. For example we might be interested in Olympic record or records in World Cricket Cup. When we are interested in successive maximum observations then records are known as *Upper Records* and when one is interested in successive minimum observations then records are known as *Lower Records*. Chandler (1952) presented the idea of records in context with monitoring of extreme weather conditions. Formally, the record time and upper record values are defined as follows.

Let $\{X_n; n = 1, 2, \ldots\}$ be a sequence of iid random variables with a continuous distribution function F. The random variables

$$L(1) = 1$$
$$L(n+1) = \min\left\{j > L(n); X_j > X_{U(n)}\right\}; n \in \mathbb{N}$$

are called the record time and $X_{U(n)}$ is called Upper Record Values. The joint density of first n upper record values is

$$f_{X_{U(1)},\ldots,X_{U(n)}}(x_1,\ldots,x_r)=\left\{\prod_{i=1}^{n-1}\frac{f(x_i)}{1-F(x_i)}\right\}f(x_n). \tag{1.4}$$

Record values have wide spread applications in reliability theory. We will discuss the upper record values in Chap. 3 and lower record values in Chap. 5.

1.2.5 k–Record Values

The upper record values provide information about largest observation in a sequence of records. Often we are interested in knowing about specific record number. The k–Record values provide basis for studying distributional behavior of such observations. The k–Record values are formally defined by Dziubdziela and Kopocinski (1976) as below.

Let $\{X_n; n=1,2,\ldots\}$ be a sequence of iid random variables with a continuous distribution function F and let k be a positive integer. The random variables $U_K(n)$ defined as $U_K(1)=1$ and

$$U_K(n+1)=\min\left\{r>U_K(n): X_{r:r+k-1}>X_{U_K(n),U_K(n)+k-1}\right\}; n\in\mathbb{N}.$$

where $X_{r:r+k-1}$ is rth order statistics based on a sample of size $r+k-1$; are called the record time and $X_{U_K(n),U_K(n)+k-1}$ is called nth $k-$record values. The joint density of n k–records is

$$f_{U_K(1),\ldots,U_K(n)}(x_1,\ldots,x_r)=k^n\left\{\prod_{i=1}^{n-1}\frac{f(x_i)}{1-F(x_i)}\right\}\times[1-F(x_n)]^{k-1}f(x_n). \tag{1.5}$$

The simple upper record values appear as special case of $k-$record values for $k=1$. The $k-$record values are discussed in Chap. 3.

1.2.6 Pfeifer's Record Values

The upper record values and $k-$upper record values are based upon the assumption that the sequence of random variables $\{X_n; n=1,2,\ldots\}$ have same distributions F. Often this assumption is very unrealistic to be kept intact and a general record model is needed. Pfeifer (1979) proposed a general model for record values when observations in a sequence are independent but are not identically distributed. The Pfeifer's record values are defined as below.

Let $\left\{X_j^{(n)}; n, j \in \mathbb{N}\right\}$ be a double sequence of independent random variables defined on some probability space with

$$P\left(X_j^{(n)}\right) = P\left(X_1^{(n)}\right); n, j \in \mathbb{N}$$

Define the inter record times as

$$\begin{aligned}\Delta_1 &= 1\\ \Delta_{n+1} &= \min\left\{j \in \mathbb{N}; X_j^{(n+1)} > X_{\Delta_n}^{(n)}\right\}; n \in \mathbb{N}.\end{aligned}$$

In this case the random variables $X_{\Delta_n}^{(n)}$ are called Pfeifer's record values.

The joint density function of n Pfeifer's record values is given as

$$f_{\Delta_1^{(1)},\ldots,\Delta_n^{(n)}}\left(x_1, \ldots, x_r\right) = \left\{\prod_{i=1}^{n-1} \frac{f_i\left(x_i\right)}{1 - F_{i+1}\left(x_i\right)}\right\} f_n\left(x_n\right); \tag{1.6}$$

where F_i is distribution function of the sequence until occurrence of ith record. If all random variables in the sequence are identically distributed then Pfeifer's record values transformed to simple upper records.

1.2.7 k_n–Records from Non-identical Distributions

The Pfeifer's record values and k–record values can be combined together to give rise to k_n–records from Non-identical distributions. Formally, the k_n–records from Non-identical distributions are defined as below.

Let $\left\{X_j^{(n)}; n, j \in \mathbb{N}\right\}$ be a double sequence of independent random variables defined on some probability space with $P\left(X_j^{(n)}\right) = P\left(X_1^{(n)}\right); n, j \in \mathbb{N}$ and let

$$X_j^{(n)} \sim X_1^{(n)} \sim F_n; n, j \in \mathbb{N}$$

Also let $(k_n; n \in \mathbb{N})$ be a sequence of positive integers. Define inter record times as

$$\begin{aligned}\Delta_1 &= 1;\\ \Delta_{n+1} &= \min\left\{j \in \mathbb{N}; X_{j:j+k_{n+1}-1}^{(n+1)} > X_{\Delta_n,\Delta_n+k_n-1}^{(n)}\right\}; n \in \mathbb{N};\end{aligned}$$

where $X_{j:j+k_{n+1}-1}$ is jth order statistics based on a sample of size $j + k_{n+1} - 1$; then the random variables

$$X_{\Delta_n,\Delta_n+k_n-1}^{(n)} = X_{\Delta_n,k_n}^{(n)}$$

are called k_n–records from Non-identical distributions.

The joint density function of r k_n–records from Non-identical distributions is given as

$$f_{\Delta^{(1)}_{1,k_1},\Delta^{(2)}_{2,k_2},\ldots,\Delta^{(r)}_{r,k_r}}\left(x_1,x_2,\ldots,x_r\right) = \left(\prod_{j=1}^{r} k_j\right)\prod_{i=1}^{r}\left[\left\{\frac{1-F_i\left(x_i\right)}{1-F_i\left(x_{i-1}\right)}\right\}^{k_i-1} \times \left\{\frac{f_i\left(x_i\right)}{1-F_{i+1}\left(x_i\right)}\right\}\right]; \tag{1.7}$$

where F_i is distribution function of the sequence until occurrence of ith record. If $k_n = 1$ for all $n \in \mathbb{N}$ then k_n–records from Non-identical distributions reduces to Pfeifer's record values.

Chapter 2
Order Statistics

2.1 Introduction

Order Statistics naturally appear in real life whenever we need to arrange observations in ascending order; say for example prices arranged from smallest to largest, scores scored by a player in last ten innings from smallest to largest and so on. The study of order statistics needs special considerations due to their natural dependence. The study of order statistics has attracted many statistician in the past. Formerly, order statistics are defined in the following.

Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution $F(x)$ and so all X_i are i.i.d. random variables having same distribution $F(x)$. The arranged sample $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ is called the *Ordered Sample* and the rth observation in the ordered sample; denoted as $X_{r:n}$ or $X_{(r)}$; is called the rth *Order Statistics.* The realized ordered sample is written as $x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{n:n}$. The distribution of rth order statistics and joint distribution of rth and sth order statistics are given below.

2.2 Joint Distribution of Order Statistics

The joint distribution of all order statistics plays an important role in deriving several special distributions of individual and group of order statistics. The joint distribution of all order statistics is easily derived from the marginal distributions of available random variables. We know that if we have a random sample of size n from a distribution function $F(x)$ as $X_1, X_2, \ldots, X_n$ then the joint distribution of all sample observations is

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i);$$

M.Q. Shahbaz et al., *Ordered Random Variables: Theory and Applications*,
Atlantis Studies in Probability and Statistics 9, DOI 10.2991/978-94-6239-225-0_2

where $f(x_i)$ is the density function of X_i. Now since all possible ordered permutations of $X_1, X_2, \ldots, X_n$ can be done in $n!$ ways, therefore the joint density function of all order statistics is readily written as

$$f(x_{1:n}, x_{2:n}, \ldots, x_{n:n}) = n! \prod_{i=1}^{n} f(x_i). \tag{2.1}$$

The joint density function of all order statistics given in (2.1) is very useful in deriving the marginal density function of a single and group of order statistics.

The joint density function of all order statistics is useful in deriving the distribution of a set of order statistics. Specifically, the joint distribution of r order statistics; $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{r:n}$; is derived as below

$$\begin{aligned}
f_{1,\ldots r:n}(x_1, \ldots, x_r) &= \int_{x_r}^{\infty} \cdots \int_{x_r}^{x_{r+3}} \int_{x_r}^{x_{r+2}} f(x_{1:n}, x_{2:n}, \ldots, x_{n:n}) \\
&\quad \times dx_{r+1} \cdots dx_n \\
&= \int_{x_r}^{\infty} \cdots \int_{x_r}^{x_{r+3}} \int_{x_r}^{x_{r+2}} n! \prod_{i=1}^{n} f(x_i) dx_{r+1} \cdots dx_n \\
&= n! \prod_{i=1}^{r} f(x_i) \int_{x_r}^{\infty} \cdots \int_{x_r}^{x_{r+3}} \int_{x_r}^{x_{r+2}} \prod_{i=r+1}^{n} f(x_i) \\
&\quad \times dx_{r+1} \cdots dx_n \\
&= \frac{n!}{(n-r)!} \left[\prod_{i=1}^{r} f(x_i) \right] [1 - F(x_r)]^{n-r}. \qquad (2.2)
\end{aligned}$$

Expression (2.2) can be used to obtain the joint marginal distribution of any specific number of order statistics.

The distribution of a single order statistics and joint distribution of two order statistics has found many applications in diverse areas of life. In the following we present the marginal distribution of a single order statistics.

2.3 Marginal Distribution of a Single Order Statistics

The marginal distribution of rth order statistics $X_{r:n}$ can be obtained in different ways. The distribution can be obtained by first obtaining the distribution function of $X_{r:n}$ and then that distribution function can be used to obtain the density function of $X_{r:n}$ as given in Arnold, Balakrishnan and Nagaraja (2008) and David and Nagaraja (2003). We obtain the distribution function of $X_{r:n}$ by first obtaining distribution function of $X_{n:n}$; the largest observation; and $X_{1:n}$; the smallest observation.

The distribution function of $X_{n:n}$ is denoted as $F_{n:n}(x)$ and is given as

$$\begin{aligned} F_{n:n}(x) &= P\{X_{n:n} \le x\} \\ &= P\{all\ X_i \le x\} = F^n(x). \end{aligned} \tag{2.3}$$

Again the distribution function of $X_{1:n}$; denoted as $F_{1:n}(x)$; is

$$\begin{aligned} F_{n:n}(x) &= P\{X_{1:n} \le x\} = 1 - P\{X_{1:n} > x\} \\ &= 1 - P\{all\ X_i > x\} = 1 - [1 - F(x)]^n. \end{aligned} \tag{2.4}$$

Now the distribution function of $X_{r:n}$; the rth order statistics; is denoted as $F_{r:n}(x)$ and is given as

$$\begin{aligned} F_{r:n}(x) &= P\{X_{r:n} \le x\} \\ &= P\{atleast\ r\ of\ X_i\ are\, less\, than\, or\, equal\, to\, x\} \\ &= \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-i}. \end{aligned} \tag{2.5}$$

Now using the relation

$$\sum_{i=r}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \int_0^p \frac{n!}{(r-1)!(n-r)!} t^{r-1}(1-t)^{n-r} dt;$$

the distribution function of $X_{r:n}$ is given as

$$\begin{aligned} F_{r:n}(x) &= \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-i} \\ &= \int_0^{F(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1}(1-t)^{n-r} dt \\ &= \int_0^{F(x)} \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} t^{r-1}(1-t)^{n-r} dt \\ &= \frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r} dt \\ &= I_{F(x)}(r, n-r+1); \end{aligned} \tag{2.6}$$

where $I_x(a, b)$ is incomplete Beta Function ratio. From (2.6) we see that the distribution function of $X_{r:n}$ resembles with the distributions proposed by Eugene, Lee and Famoye (2002). Expression (2.6) is valid either if sample has been drawn from a discrete distribution. An alternative form for the distribution function of $X_{r:n}$ is given as

$$F_{r:n}(x) = \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-i}$$
$$= \sum_{i=r}^{n} \binom{n}{i} F^i(x) \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k} F^k(x)$$
$$= \sum_{i=r}^{n} \sum_{k=0}^{n-i} (-1)^k \binom{n}{i} \binom{n-i}{k} F^{i+k}(x). \quad (2.7)$$

Assuming that $X_i^{\prime}s$ are absolutely continuous, the density function of $X_{r:n}$; denoted by $f_{r:n}(x)$; is easily obtained from (2.6) as below

$$f_{r:n}(x) = \frac{d}{dx} F_{r:n}(x)$$
$$= \frac{d}{dx}\left[\frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r} dt\right]$$
$$= \frac{1}{B(r, n-r+1)}\left[\frac{d}{dx}\left\{\int_0^{F(x)} t^{r-1}(1-t)^{n-r} dt\right\}\right]$$
$$= \frac{1}{B(r, n-r+1)} f(x) F^{r-1}(x)[1 - F(x)]^{n-r}$$
$$= \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x)[1 - F(x)]^{n-r}. \quad (2.8)$$

The density function of $X_{1:n}$ and $X_{n:n}$ can be immediately written from (2.8) as

$$f_{1:n}(x) = nf(x)[1 - F(x)]^{n-1}$$

and

$$f_{n:n}(x) = nf(x)F^{n-1}(x).$$

The distribution of $X_{r:n}$ can also be derived by using the multinominal distribution as under:

Recall that probability mass function of multinomial distribution is

$$P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k) = \frac{n!}{y_1! y_2! \ldots y_k!} p_1^{y_1} p_2^{y_2} \ldots p_k^{y_k};$$

and can be used to compute the probabilities of joint occurrence of events. Now the place of $x_{r;n}$ in ordered sample can be given as

$$\underbrace{x_{1:n} \le x_{2:n} \le \ldots \le x_{r-1:n}}_{\substack{r-1 \; observations \\ Event \; 1}} \le \underbrace{x_{r:n}}_{Event \; 2} \le \underbrace{x_{r+1:n} \le \ldots \le x_{n:n}}_{\substack{n-r \; observations \\ Event \; 3}}$$

In the above probability of occurrence of *Event 1* is $F(x)$, that of *Event 2* is $f(x)$ and probability of *Event 3* is $[1 - F(x)]$. Hence the joint occurrence of above three events; which is equal to density of $X_{r:n}$ is

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x)[1 - F(x)]^{n-r};$$

which is (2.8). When the density $f(x)$ is symmetrical about μ then the distributions of rth and $(n - r + 1)th$ order statistics are related by relation

$$f_{r:n}(\mu + x) = f_{n-r+1:n}(\mu - x).$$

Above relation is very useful in moment relations of order statistics.

When the sample has been drawn from a discrete distribution with distribution function $F(x)$ then the density of $X_{r:n}$ can be obtained as below

$$\begin{aligned} f_{r:n}(x) &= F_{r:n}(x) - F_{r:n}(x-1) \\ &= I_{F(x)}(r, n-r+1) - I_{F(x-1)}(r, n-r+1) \\ &= P\big\{F(x-1) < T_{r,n-r+1} < F(x)\big\} \\ &= \frac{1}{B(r, n-r+1)} \int_{F(x-1)}^{F(x)} u^{r-1}(1-u)^{n-r} du. \end{aligned} \tag{2.9}$$

Expression (2.9) is the probability mass function of $X_{r:n}$ when sample is available from a discrete distribution. The probability mass function of $X_{r:n}$ can also be written in binomial sum as under

$$\begin{aligned} f_{r:n}(x) &= F_{r:n}(x) - F_{r:n}(x-1) \\ &= \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-i} \\ &\quad - \sum_{i=r}^{n} \binom{n}{i} F^i(x-1)[1 - F(x-1)]^{n-i} \\ &= \sum_{i=r}^{n} \binom{n}{i} \big\{F^i(x)[1 - F(x)]^{n-i} \\ &\quad - F^i(x-1)[1 - F(x-1)]^{n-i}\big\}. \end{aligned} \tag{2.10}$$

We now obtain the joint distribution of two ordered observations, namely $X_{r:n}$ and $X_{s:n}$ for $r \le s$; in the following.

2.4 Joint Distribution of Two Order Statistics

Suppose we have random sample of size n from $F(x)$ and observations are arranged as $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$. The joint distribution function of $X_{r:n}$ and $X_{s:n}$ for $r \leq s$ is given by Arnold et al. (2008) as

$$\begin{aligned} F_{r,s:n}(x_r, x_s) &= P(X_{r:n} \leq x_r, X_{s:n} \leq x_s) \\ &= P(atleast\ r\ of\ X_i\ are\ less\ than\ or\ equal\ to\ x_r \\ &\quad and\ atleast\ s\ of\ X_i\ are\ less\ than\ or\ equal\ to\ x_s) \\ &= \sum_{j=s}^{n}\sum_{i=r}^{s} P(Exactly\ r\ of\ X_i\ are\ less\ than\ or\ equal\ to\ x_r \\ &\quad and\ exactly\ s\ of\ X_i\ are\ less\ than\ or\ equal\ to\ x_s) \\ &= \sum_{j=s}^{n}\sum_{i=r}^{s} \frac{n!}{i!(j-i)!(n-j)!} F^i(x_r) \\ &\quad \times [F(x_s) - F(x_r)]^{j-i}[1 - F(x_s)]^{n-j}. \end{aligned}$$

Now using the relation

$$\sum_{j=s}^{n}\sum_{i=r}^{s} \frac{n!}{i!(j-i)!(n-j)!} p_1^i (p_2 - p_1)^{j-i}(1 - p_2)^{n-j}$$
$$= \int_0^{p_1}\int_{t_1}^{p_2} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} t_1^{r-1}(t_2 - t_1)^{s-r-1}(1 - t_2)^{n-s} dt_2 dt_1;$$

we can write the joint distribution function of two order statistics as

$$\begin{aligned} F_{r,s:n}(x_r, x_s) &= \int_0^{F(x_r)}\int_{t_1}^{F(x_s)} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \\ &\quad \times t_1^{r-1}(t_2 - t_1)^{s-r-1}(1 - t_2)^{n-s} dt_2 dt_1 \qquad (2.11) \\ &; -\infty < x_r < x_s < \infty; \end{aligned}$$

which is incomplete bivariate beta function ratio. Expression (2.11) holds for both discrete and continuous random variables. When $F(x)$ is absolutely continuous then density function of $X_{r:n}$ and $X_{s:n}$ can be obtained from (2.11) and is given as

$$\begin{aligned} f_{r,s:n}(x_r, x_s) &= \frac{d^2}{dx_r dx_s} F_{r,s:n}(x_r, x_s) \\ &= \frac{d^2}{dx_r dx_s}\Big[\int_0^{F(x_r)}\int_{t_1}^{F(x_s)} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \\ &\quad t_1^{r-1}(t_2 - t_1)^{s-r-1}(1 - t_2)^{n-s} dt_2 dt_1\Big] \end{aligned}$$

or

$$\begin{aligned}
f_{r,s:n}(x_r, x_s) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \times \\
&\quad \frac{d^2}{dx_r dx_s}\left[\int_0^{F(x_r)} \int_{t_1}^{F(x_s)} t_1^{r-1}(t_2-t_1)^{s-r-1}(1-t_2)^{n-s} dt_2 dt_1\right] \\
&= \frac{1}{B(r, s-r, n-s+1)} f(x_r) f(x_s) F^{r-1}(x_r) \\
&\quad \times \left[F(x_s) - F(x_r)\right]^{s-r-1} [1-F(x_s)]^{n-s}
\end{aligned}$$

or

$$\begin{aligned}
f_{r,s:n}(x_r, x_s) &= C_{r,s,n} f(x_r) f(x_s) F^{r-1}(x_r)[F(x_s) - F(x_r)]^{s-r-1} \\
&\quad \times [1-F(x_s)]^{n-s}, \; -\infty < x_r < x_s < \infty, \qquad (2.12)
\end{aligned}$$

where $C_{r,s,n} = [B((r, s-r, n-s+1))]^{-1} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$.

The joint probability mass function $P\{X_{r:n} = x; X_{s:n} = y\}$ of $X_{r:n}$ and $X_{s:n}$ can be obtained by using the fact that

$$\begin{aligned}
f_{r,s:n}(x, y) &= F_{r,s:n}(x, y) - F_{r,s:n}(x-1, y) \\
&\quad - F_{r,s:n}(x, y-1) + F_{r,s:n}(x-1, y-1) \\
&= P\{F(x-1) < T_r \le F(x), F(y-1) < T_s \le F(y)\} \\
&= C_{r,s,n} \int\int_B v^{r-1}(w-v)^{s-r-1}(1-w)^{n-s} dv dw \; ; \qquad (2.13)
\end{aligned}$$

where integration is over the region

$$\{(v, w) : v \le w, F(x-1) \le v \le F(x), F(y-1) \le w \le F(y)\}\cdot$$

Consider the joint distribution of two order statistics as

$$\begin{aligned}
f_{r,s:n}(x_r, x_s) &= C_{r,s,n} f(x_r) f(x_s) F^{r-1}(x_r)\left[F(x_s) - F(x_r)\right]^{s-r-1} \\
&\quad \times [1-F(x_s)]^{n-s};
\end{aligned}$$

where $C_{r,s,n} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\cdot$ Using $r = 1$ and $s = n$ the joint density of smallest and largest observation is readily written as

$$f_{1,n:n}(x_1, x_n) = n(n-1) f(x_1) f(x_n)[F(x_n) - F(x_1)]^{n-2}. \qquad (2.14)$$

Further, for $s = r + 1$ the joint distribution of two contiguous order statistics is

$$f_{r,r+1:n}(x_r, x_{r+1}) = \frac{n!}{(r-1)!(n-r-1)!} f(x_r) f(x_{r+1}) F^{r-1}(x_r) \times \Big[1 - F(x_{r+1})\Big]^{n-r-1}. \quad (2.15)$$

Analogously, the joint distribution of any k order statistics $X_{r_1:n}, X_{r_2:n}, \ldots, X_{r_k:n}$; for $x_1 \le x_2 \le \cdots \le x_k$; is

$$f_{r_1,r_2,\ldots,r_k:n}(x_1, x_2, \cdots, x_k) = n! \prod_{j=0}^{k} \left\{ \frac{\big[F(x_{r+1}) - F(x_r)\big]^{r_{j+1}-r_j-1}}{\big(r_{j+1} - r_j - 1\big)!} \right\} \times \left\{ \prod_{j=1}^{k} f\big(x_j\big) \right\}. \quad (2.16)$$

where $x_0 = -\infty, x_{n+1} = +\infty, r_0 = 0$ and $r_{n+1} = n + 1$. Expression (2.16) can be used to obtain joint distribution of any number of ordered observations.

Example 2.1 A random sample is drawn from Uniform distribution over the interval [0, 1]. Obtain distribution of rth order statistics and joint distribution of two order statistics.

Solution: The density and distribution function of $U(0, 1)$ are

$$f(u) = 1; \;\; F(u) = u.$$

The distribution of rth order statistics is

$$f_{r:n}(u) = \frac{1}{B(r, n-r+1)} f(u) F^{r-1}(u)[1 - F(u)]^{n-r} = \frac{1}{B(r, n-r+1)} u^{r-1}(1-u)^{n-r};$$

which is a Beta random variable with parameters r and $n - r + 1$. Again the joint distribution of two order statistics is

$$f_{r,s:n}(u_r, u_s) = \frac{1}{B(r, s-r, n-s+1)} f(u_r) f(u_s) F^{r-1}(u_r) \times \Big[F(u_s) - F(u_r)\Big]^{s-r-1} [1 - F(u_s)]^{n-s} = \frac{1}{B(r, s-r, n-s+1)} u_r^{r-1}(u_s - u_r)^{s-r-1}(1-u_s)^{n-s}.$$

The joint distribution of largest and smallest observation is immediately written as

$$f_{r,s:n}(u_r, u_s) = n(n-1)(u_n - u_1)^{n-2}.$$

Example 2.2 A random sample of size n is drawn from the standard power function distribution with density

$$f(x) = vx^{v-1}; 0 < x < 1, v > 0.$$

Obtain the distribution of rth order statistics and joint distribution of rth and sth statistics.

Solution: For given distribution we have

$$F(x) = \int_0^x f(t)dt = \int_0^x vt^{v-1}dt = x^v; 0 < x < 1.$$

Now distribution of rth order statistics is

$$\begin{aligned} f_{r:n}(x) &= \frac{1}{B(r, n-r+1)} f(x)F^{r-1}(x)[1-F(x)]^{n-r} \\ &= \frac{1}{B(r, n-r+1)} vx^{rv-1}(1-x^v)^{n-r}. \end{aligned}$$

The distribution function of rth order statistics is readily written as

$$\begin{aligned} F_{r:n}(x) &= \frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}dt \\ &= \frac{1}{B(r, n-r+1)} \int_0^{x^v} t^{r-1}(1-t)^{n-r}dt \\ &= \sum_{i=r}^{n} \binom{n}{i} x^{iv}(1-x^v)^{n-i}. \end{aligned}$$

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

$$\begin{aligned} f_{r,s:n}(x_r, x_s) &= \frac{1}{B(r, s-r, n-s+1)} f(x_r)f(x_s)F^{r-1}(x_r) \\ &\quad \times \left[F(x_s) - F(x_r) \right]^{s-r-1} [1-F(x_s)]^{n-s} \\ &= C_{r,s,n} v^2 x_r^{rv-1} x_s^{v-1} \left(x_s^v - x_r^v\right)^{s-r-1} \left(1 - x_s^v\right)^{n-s}. \end{aligned}$$

where $C_{r,s,n} = [B(r, s-r, n-s+1)]^{-1} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$.

2.5 Distribution of Range and Other Measures

Suppose a random sample of size n is available from $F(x)$ and let $X_{r:n}$ be the rth order statistics. Further let $X_{s:n}$ be sth order statistics with $r < s$. The joint density function of $X_{r:n}$ and $X_{s:n}$ is

$$f_{r,s:n}(x_r, x_s) = C_{r,s,n} f(x_r) f(x_s) F^{r-1}(x_r) \Big[F(x_s) - F(x_r) \Big]^{s-r-1} \\ \times [1 - F(x_s)]^{n-s}.$$

Using above we can obtain the density of $W_{rs} = X_{s:n} - X_{r:n}$ by making the transformation $w_{rs} = x_s - x_r$. The joint density of w_{rs} and x_r in this case is

$$f_{W_{rs}}(w_{rs}) = C_{r,s,n} f(x_r) f(x_r + w_{rs}) F^{r-1}(x_r) \\ \times [F(x_r + w_{rs}) - F(x_r)]^{s-r-1} [1 - F(x_r + w_{rs})]^{n-s}.$$

The marginal density of w_{rs} is

$$f_{W_{rs}}(w_{rs}) = C_{r,s,n} \int_{-\infty}^{\infty} f(x_r) f(x_r + w_{rs}) F^{r-1}(x_r) \\ \times [F(x_r + w_{rs}) - F(x_r)]^{s-r-1} [1 - F(x_r + w_{rs})]^{n-s} dx_r.$$

When $r = 1$ and $s = n$ then above result provide the density function of *Range* (w) in a sample of size n and is given as

$$f_W(w) = n(n-1) \int_{-\infty}^{\infty} f(x_r) f(x_r + w) [F(x_r + w) - F(x_r)]^{n-2} dx_r. \quad (2.17)$$

The distribution function of sample range can be easily obtained from (2.17) and is

$$\begin{aligned} F_W(w) &= n \int_{-\infty}^{\infty} f(x_r) \int_0^w (n-1) f\big(x_r + w^/\big) \\ &\quad \times \big[F\big(x_r + w^/\big) - F(x)\big]^{n-2} dw^/ dx_r \\ &= n \int_{-\infty}^{\infty} f(x_r) \big[F\big(x_r + w^/\big) - F(x_r)\big]^{n-1} \Big|_{w^/=0}^{w^/=w} dx_r \\ &= n \int_{-\infty}^{\infty} f(x_r) [F(x_r + w) - F(x_r)]^{n-1} dx_r. \quad (2.18) \end{aligned}$$

Again suppose that number of observations in sample are even; say $n = 2m$; then we know that the sample median is

$$\widetilde{X} = \frac{1}{2}\big[X_{m:n} + X_{m+1:n}\big].$$

The distribution of median can be obtained by using joint distribution of two contiguous order statistics and is given as

$$f_{m,m+1:n}(x_m, x_{m+1}) = C^{/}_{m,n} f(x_m) f(x_{m+1}) F^{m-1}(x_m)\big[1 - F(x_{m+1})\big]^{m-1};$$

where $C^{/}_{m,n} = \frac{n!}{[(m-1)!]^2}\cdot$ Now making the transformation $\widetilde{x} = \frac{1}{2}\big[x_m + x_{m+1}\big]$ and $y = x_m$ the jaccobian of transformation is 2 and hence the joint density of $\widetilde{x}$ and y is

$$f_{\widetilde{X}Y}(\widetilde{x}, y) = 2C^{/}_{m,n} f(y) f(2\widetilde{x} - y) F^{m-1}(y)[1 - F(2\widetilde{x} - y)]^{m-1}.$$

The marginal density of sample median is, therefore

$$f_{\widetilde{X}}(\widetilde{x}) = 2C^{/}_{m,n} \int_{-\infty}^{\widetilde{x}} f(y) f(2\widetilde{x} - y) F^{m-1}(y)[1 - F(2\widetilde{x} - y)]^{m-1} dy \qquad (2.19)$$

The density of sample median for an odd sample size; say $n = 2m + 1$; is simply the density of mth order statistics for a sample of size $2m + 1$.

Example 2.3 Obtain the density function of sample range for a sample of size n from uniform distribution with density

$$f(x) = 1; 0 < x < 1.$$

Solution: The distribution of sample range for a sample of size n from distribution $F(x)$ is

$$f_W(w) = n(n-1) \int_{-\infty}^{\infty} f(x_r) f(x_r + w)[F(x_r + w) - F(x_r)]^{n-2} dx_r.$$

Now for uniform distribution we have

$$f(x) = 1;\ F(x) = x.$$

So

$$f(x_r + w) = 1;\ F(x_r + w) = (x_r + w),$$

hence the density function of range is

$$\begin{aligned} f_W(w) &= n(n-1) \int_0^{1-w} [(x_r + w) - x_r]^{n-2} dx_r \\ &= n(n-1) \int_0^{1-w} w^{n-2} dx_r \\ &= n(n-1) w^{n-2}(1 - w); 0 < w < 1. \end{aligned}$$

Example 2.4 Obtain the density function of range in a sample of size 3 from exponential distribution with density

$$f(x) = e^{-x}; x > 0.$$

Solution: The distribution of sample range for a sample of size n from distribution $F(x)$ is

$$f_W(w) = n(n-1)\int_{-\infty}^{\infty} f(x_r)f(x_r+w)[F(x_r+w)-F(x_r)]^{n-2}dx_r.$$

which for $n = 3$ becomes

$$f_W(w) = 6\int_{-\infty}^{\infty} f(x_r)f(x_r+w)[F(x_r+w)-F(x_r)]dx_r.$$

Now for exponential distribution we have

$$f(x) = e^{-x} \text{ and } F(x) = 1 - e^{-x},$$

hence

$$f(x_r) = e^{-x_r} \text{ and } F(x_r) = 1 - e^{-x_r},$$
$$f(w+x_r) = e^{-(w+x_r)} \text{ and } F(w+x_r) = 1 - e^{-(w+x_r)}.$$

Using these in above expression we have

$$\begin{aligned} f_W(w) &= 6\int_0^{\infty} e^{-x_r}e^{-(w+x_r)}\left[e^{-x_r} - e^{-(w+x_r)}\right]dx_r \\ &= 6e^{-w}\left(1-e^{-w}\right)\int_0^{\infty} e^{-3x_r}dx_r \\ &= 2e^{-w}\left(1-e^{-w}\right); w > 0, \end{aligned}$$

as required density function of range.

Example 2.5 Obtain the density function of median in a sample of size $n = 2m$ from exponential distribution with density function

$$f(x) = e^{-x}; x > 0.$$

Solution: The distribution of sample median for a sample of size $n = 2m$ from distribution $F(x)$ is

$$f_{\widetilde{X}}(\widetilde{x}) = 2C^{/}_{m,n}\int_{-\infty}^{\widetilde{x}} f(y)f(2\widetilde{x}-y)F^{m-1}(y)[1-F(2\widetilde{x}-y)]^{m-1}dy;$$