Faculty of Education
Mathematics Department

STUDY ON ORDERED OBSERVATIONS AND STATISTICAL INFERENCES

A Thesis

Submitted in Partial Fulfillment of the Requirements of the Master's Degree in Teacher's Preparation in Science

(Mathematical Statistics)

Submitted to:

The Department of Mathematics, Faculty of Education, Ain Shams University

By

Mohamed Said Mohamed Mahmoud

Demonstrator at,
Mathematics Department, Faculty of Education, Ain Shams University

Supervised by

Prof. Dr. M. M. Mohie El-Din
Professor of Mathematical Statistics
Faculty of Science
Al-Azhar University

Dr. Nahed Said Abd El-Latef Ali
Lecturer of Mathematical Statistics
Faculty of Education
Ain Shams University

Dr. Montaser Mahmoud Amein Mohamed
Lecturer of Mathematical Statistics
Faculty of Science
Al-Azhar University

(2013)

Faculty of Education
Mathematics Department

Candidate: Mohamed Said Mohamed Mahmoud

Thesis Title: STUDY ON ORDERED OBSERVATIONS AND STATISTICAL INFERENCES

Degree: Master for Teacher's Preparation in Science

(Mathematical Statistics)

Supervisors:

No.	Name	Profession	Signature
1.	**Prof. Dr. M. M. Mohie El-Din**	Professor of Mathematical Statistics, Mathematics Department, Faculty of Science, Al-Azhar University.	
2.	**Dr. Nahed Said Abd El-Latef Ali**	Lecturer of Mathematical Statistics, Mathematics Department, Faculty of Education, Ain Shams University.	
3.	**Dr. Montaser Mahmoud Amein Mohamed**	Lecturer of Mathematical Statistics, Mathematics Department, Faculty of Science, Al-Azhar University.	

Acknowledgements

First of all gratitude and thanks to gracious **Allah** who always helps and guides me. I would like to thank **the prophet Mohamed** "peace be upon him" who urges us to seek knowledge and who is the teacher of mankind.

I would like also to thank the supervision committee who are:

Prof. Dr. M. M. Mohie El-Din, Professor of Mathematical Statistics, Faculty of Science, Al-Azhar University, for accepting supervision of me, learning me ethics and scientific research assets, his good care for me and my colleagues and for his support, encouragement and continuing guidance during preparing this thesis. He discussed with me many research problems in the seminar of statistics which was held in Faculty of Science, Al-Azhar University.

Dr. Nahed Said Abd El-Latef Ali, Lecturer of Mathematical Statistics, Faculty of Education, Ain Shams University, who provided me with guidance and continuous encouragement. She offered me much of her precious time and provided me with her wisdom and knowledge through many discussions we had.

Dr. Montaser Mahmoud Amein Mohamad, Lecturer of Mathematical Statistics, Faculty of Science, Al-Azhar University, who helped me at the first step in this study through his suggestions for the research problems, valuable instructions, guidance and continuous follow up in this study. He offered me much of his precious time and provided me with his wisdom and knowledge through many discussions we had.

Thanks also are due to **Dr. Raafat Riad Rizk allah**, Head of Mathematics Department, Faculty of Education, Ain Shams University, and all staff members for providing me with all facilities required to the success of this work.

Thanks also are due to **Prof. Dr. Galal Mahrous**, Vice Dean for Student Affairs, Faculty of Education, Ain Shams University, for his support and continuous encouragement.

Many thanks to all my teachers and my fellow companions of scientific research in the school of mathematical statistics, headed by Prof. Dr. M. M. Mohie El-Din.

Finally, I am appreciative to my kind parents, my wife, and my beloved family for their support, patience, sacrifice and continuous encouragement. Finally gratitude and thanks to gracious **Allah** who always helps and guides me.

Mohamed Said Mohamed Mahmoud

Contents

Abbreviations and symbols

1. pdf: Probability density function.
2. cdf: Cumulative distribution function.
3. iid: Independent identically distributed.
4. MLE: Maximum Likelihood Estimation.
5. $CRLB$: The Cramer-Rao Lower Bound.
6. $UMVUE$: Uniformly Minimum Variance Unbiased Estimation.
7. MSE: Mean Square Error.
8. HPD: Highest Posterior Density.
9. $AMLE$: Approximate Maximum Likelihood Estimation.
10. $ABLUE$: Approximate Best Linear Unbiased Estimation.
11. EFF: Relative Efficiency.
12. EM algorithm: Expectation Maximization algorithm.
13. $X_*^{(r)}$: The r-th sequential order statistics.
14. $Exp(1)$: Standard exponential distribution.
15. $\Gamma(a, b)$: Gamma distributed random variable.
16. $IG(a, b)$: Inverted gamma distributed random variable.

17. $I_A(.)$: Indicator function of A.

18. $u(0,1)$: Standard uniform distribution.

19. π^{θ}: Prior density function.

20. $\pi^{\theta|X}(.|x)$: Posterior density function.

21. $\mu^{(k)} = E(X^k)$: The k-th moment about the origin of the random variable X.

22. $X(r,n,\tilde{m},k)$: The r-th generalized order statistics.

Summary

Ordered observations appear in many parts of statistics and play an important role in applied statistics. The first model appear in the ordered data is the order statistics (ordinary order statistics) and record values which are widely used in statistical modeling and inference, both models describe random variables arranged in order of magnitude. In addition to these well known models, their are several other models of ordered random variables like: sequential order statistics, k-th record values, Pfeifer's record model, etc., which are naturally suggested by a statistical application in reliability theory and life testing. In the distribution theoretical sense, all of these models of ordered random variables are contained in the proposed concept of generalized order statistics which has been introduced by Kamps ([27], [28]).

The notion of sequential order statistics has been introduced in Kamps [27] in order to model effects of component failures on the reliability of k-out-of-n systems. The structures of k-out-of-n often appear in technical systems and play an important role in reliability theory and life testing.

For that, we choose the subject of the thesis "Study on Ordered Observations and Statistical Inferences" aiming at a comprehensive overview of some statistical properties of ordered random variables such as sequential order statistics and generalized order statistics.

The thesis consists of four chapters:

Chapter 1

This chapter is considered as an introductory chapter. It contains definitions, concepts and theorems. A summary of the previous studies is introduced such as: distribution of an order statistics, types of ordered data, sequential order statistics, generalized order statistics, k-th record values and mixture distributions.

Chapter 2

In this chapter, some statistical inference techniques including MLE and UMVUE with Type-II censored sequential order statistics from a hierarchy of Pareto models are obtained. Also, Bayes estimation, HPD and Bayes prediction with Multiply Type-II censored sequential order statistics from Pareto(I) and Weibull distributions are obtained. Some results of this chapter are published in the journal of Advanced Research in Applied Mathematics, Volume 5, Issue 3, April 2013.

Chapter 3

Our goal in this chapter is to extend the approximate maximum likelihood estimators (Jones's method) to estimate the location parameter (mode) with generalized order statistics for Weibull distribution. Some special cases including sequential order statistics and k-th record values are obtained. Finally, some numerical computations and comparison between AMLEs and ABLUEs are given. Some results of this chapter are put in the form of research for publication.

Chapter 4

Finally in Chapter 4, we can't deal with the MLE in many situations to estimate the parameters especially in mixture distributions, so the EM algorithm technique is used to estimate the parameters of finite mixture models with sequential order statistics. Some results of this chapter are accepted for publication in the International Organization of Scientific Research (Journal of Mathematics).

Chapter 1

Introduction

The introductory chapter is considered as a background for the material included in the thesis. The purpose of this chapter is to present a short survey of some needed definitions and theories of the material used in this thesis.

1.1 Order statistics

Order statistics appear in many parts of statistics and play an important role in applied statistics. They and their moments gained their importance in many statistical problems. They are also more applicable in many engineering fields since in these cases the smallest or the largest future realization of a random variable is important than the mean or the median of the distribution. For example, the lowest strength value of critical structural component is an important factor in assessing the value of structural design and the largest value is also important for similar assessment.

Definition 1.1.1 *Let $X_1, X_2, ..., X_n$ be a random sample from an absolutely continuous population with pdf $f(x)$ and cdf $F(x)$, $X_i's$ are arranged in nondecreasing order. Then the smallest of the $X_i's$ is denoted by $X_{1:n}$, the second smallest is denoted by $X_{2:n}$,..., and, finally,* *the largest is denoted by $X_{n:n}$. Thus $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ are called the order statistics obtained by arranging the preceding random sample in increasing order of magnitude, see Arnold et al. [4].*

1.1.1 The probability density functions of order statistics

The joint *pdf* of the order statistics $X_{1:n}, X_{2:n}, ..., X_{n:n}$ is given by:

$$f_{1,2,...,n:n}(x_1, x_2, ..., x_n) = n! \prod_{r=1}^{n} f(x_r), \ -\infty < x_1 \leq x_2 \leq ... \leq x_n < \infty. \tag{1.1}$$

The *pdf* of r-th order statistic $(1 \leq r \leq n)$ in a sample of size n is given by:

$$f_{r:n}(x) = C_{r:n}[F(x)]^{r-1}[1 - F(x)]^{n-r} f(x), \ -\infty < x < \infty, \tag{1.2}$$

where $f(.)$ and $F(.)$ are the *pdf* and *cdf* of the random variable X and

$$C_{r:n} = \frac{n!}{(r-1)!(n-r)!} \ .$$

The joint *pdf* of $X_{r:n}$ and $X_{s:n}$ where $(1 \leq r < s \leq n)$ is given by:

$$f_{r,s:n}(x, y) = C_{r,s:n}[F(x)]^{r-1}[F(y) - F(x)]^{s-r-1}[1 - F(y)]^{n-s} f(x) f(y), \quad -\infty < x < y < \infty, \tag{1.3}$$

where

$$C_{r,s:n} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}.$$

For more details about the probability density functions of order statistics, see Arnold et al. [4].

1.1.2 Types of ordered censored data

Let us consider a life testing experiment where n items are kept under observation until failure. These items could be some systems,

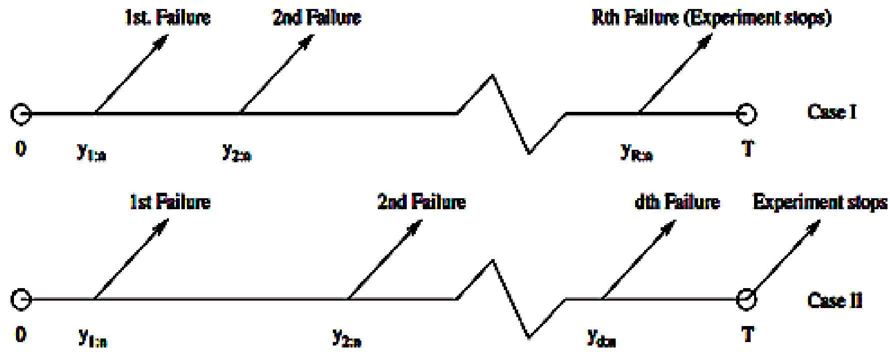

Figure 1.1: *Censoring schemes.*

components, or computer chips in reliability study experiments, or they could be patients put under certain drug or clinical conditions. Let the life lengths of these n items be *iid* random variables with a common absolutely continuous *cdf* $F(x;\theta)$ and *pdf* $f(x;\theta)$ where θ is the unknown parameter, then we have a random sample $X_1, X_2, ..., X_n$ from the *cdf* $F(x;\theta)$.

It should be noted that, these values are recorded in increasing order of magnitude; that is, the data appear as the vector of order statistics in a natural way. Thus, the life length is the random variable which describe the length, number, time or any thing else that the item takes until it fails or the whole system fails. The ordered censored data can be shown as follows:

1. Suppose that the experiment is terminated at a predetermined time T. Only the failure times of the items that failed prior to this time are recorded. The data so obtained constitute a Type-I censored data (see Figure (1.1) caseII).

2. Type-II censored data is obtained when the experiment is terminated at the R-th failure. Here R is fixed and the duration