Faculty of Education
Mathematics Department

STATISTICAL INFERENCE FOR SOME CONTINUOUS DISTRIBUTIONS BASED ON RANKED SET SAMPLING

A Thesis

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

(Statistics)

Submitted to:

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

By

Haidy Ali El-Sayed Newer

Assistant Lecturer at,
Mathematics Department, Faculty of Education, Ain Shams University

Supervised by

Prof. Mostafa Mohamed Mohie El-Din
Professor of Mathematical Statistics
Faculty of Science
Al-Azhar University

Dr. Ehab F. Abd-elfattah
Associate Professor of Mathematical Statistics
Faculty of Education
Ain Shams University

Dr. Mohammed Sayed Kotb
Assistant Professor of Mathematical Statistics
Faculty of Science
Al-Azhar University

(2017)

Faculty of Education
Mathematics Department

Candidate: Haidy Ali El-Sayed Newer

Thesis Title: STATISTICAL INFERENCE FOR SOME CONTINUOUS DISTRIBUTIONS BASED ON RANKED SET SAMPLING

Degree: Doctor's Philosophy of Teacher's Preparation in Science

(Statistics)

Supervisors:

No.	Name	Profession	Signature
1.	**Prof. Mostafa Mohamed Mohie El-Din**	Professor of Mathematical Statistics, Mathematics Department, Faculty of Science, Al-Azhar University.	
2.	**Dr. Ehab F. Abd-elfattah**	Associate Professor of Mathematical Statistics, Mathematics Department, Faculty of Education, Ain Shams University.	
3.	**Dr. Mohammed Sayed Kotb**	Assistant Professor of Mathematical Statistics, Mathematics Department, Faculty of Science, Al-Azhar University.	

Acknowledgements

First of all gratitude and thanks to gracious and merciful **Allah** who always helps and support me. I would like to thank **the prophet Mohamed** "peace be upon him" who is the teacher of mankind. I would like also to extend my deepest gratitude to the supervision committee who are:

Prof. Mostafa Mohamed Mohie El-Din, Professor of Mathematical Statistics, Faculty of Science, Al-Azhar University, for his excellent guidance, helpful discussions and continuous encouragement during the development of this work. Many thanks to all my colleagues research fellow at the School of Mathematical Statistics, headed by Prof. Mostafa Mohamed Mohie El-Din.

Dr. Ehab F. Abd-elfattah, Associate Professor of Mathematical Statistics, Faculty of Education, Ain Shams University, who helped me at the first step in this study through his suggestions for the research problems.

Dr. Mohammed Sayed Kotb, Assistant Professor of Mathematical Statistics, Faculty of Science, Al-Azhar University, for his suggestions, valuable instructions and continuous follow me in this work.

Thanks also are due to **Dr. Osman Mohamed Fraig**, 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 dissertation.

I extend also my heartfelt thanks to scientific research fellows at the school of Mathematical Statistics headed by **Prof. Mostafa Mohamed Mohie El-Din**.

I extend my heartfelt thanks to **Prof. Galal Mahrous**, Professor of Applied Mathematics, Faculty of Education, Ain Shams University, for his support and continuous encouragement.

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

Haidy Ali Newer

Contents

List of Abbreviations and Symbols

1. ALF: Al-Bayyati loss function.
2. BLUEs: Best linear unbiased estimators.
3. BPI: Baysian prediction interval.
4. cdf: Cumulative distribution function.
5. iid: Independent identically distributed.
6. INID: Independent non identically distributed.
7. IW : Inverse Weibull.
8. Kum-G: Kumaraswamy generalized.
9. Kum-LGIW: Kumaraswamy-log-generalized inverse Weibull.
10. L: Lower bound.
11. LExp: Linear exponential.
12. LF: Likelihood function.
13. LGIW: Log-generalized inverse Weibull.
14. LINEX: Linear expontial loss function.
15. MLE: Maximum likelihood estimation.
16. MPEL: Minimum posterior expected loss.

17. MSE: Mean squared error.
18. OS: Order statistic.
19. ORSS: Ordered ranked set sampling.
20. ORVs: Ordinary record values.
21. pdf: Probability density function.
22. Ray(α): Rayleigh distribution.
23. RRSS: Record ranked set sampling.
24. RSS: Ranked set sampling.
25. r.v's: Random variables.
26. RVs: Record values.
27. SEL: Squared error loss.
28. SRS: Simple random sample.
29. U: Upper bound.
30. $U(0,1)$: Standard uniform distribution.
31. $(\cdot)_{AL}$: Baysian estimators based on ALF.
32. $(\cdot)_{BL}$: Baysian estimators based on LINEX.
33. $(\cdot)_{BS}$: Baysian estimators based on SEL.
34. $\Gamma(a,b)$: Gamma function.
35. $\overline{F}(.)$: Survival function.

Summary

McIntyre [54] proposed ranked set sampling (RSS) as a sampling method that improve the precision of the sample mean estimator of the population mean without the bias of researcher choice and referred to it as a method of unbiased selective sampling using ranked sets. Subsequently some properties of RSS estimator of population mean such as unbiasedness, variance and relative precision with respect to simple random sampling (SRS) have been established by Takahasi and Wakimoto [68].

The aim of this thesis is to describe the structural method for obtaining RSS and study the statistical inference for some continuous distribution based on the two sampling methods; RSS and SRS.

This thesis consists of six chapters:

Chapter 1

This chapter is an introductory chapter. It consists of definitions and basic concepts which will be used in this thesis. At the end of this chapter, a literature review of the previous studies is presented.

Chapter 2

In this chapter, we provide Bayesian estimation for the parameters of the Pareto distribution based on SRS and RSS. Posterior risk function of the derived estimators are also obtained by using squared error loss (SEL) function. Two-sample Bayesian prediction for future observations are obtained by using SRS and RSS. Lastly, a simulation study is conducted to assess the performance of the proposed estimation and prediction techniques. The results of this chapter were published at:

" Journal of Statistics Applications & Probability, 2015, 4 (2), 1–11."

Chapter 3

Chapter 3 presents order statistics of independent and non identically distributed (INID) random variables to obtain ordered ranked set sampling (ORSS) under Type-II censoring scheme. Bayesian inference of unknown parameters under a SEL function of the Pareto distribution are determined. We compute posterior risk function of the derived estimators based on ORSS and compare them with those based on the corresponding SRS to assess the efficiency of the obtained estimators. Two-sample Bayesian prediction for future observations are introduced by using SRS and ORSS. A simulation study and real data are applied to show the accuracy of the proposed results. The results of this chapter were published at:

"Communications in Statistics Theory and Methods, 2017, 46 (13), 6264–6279."

Chapter 4

The aim of this chapter is to use RSS to develop Bayesian analysis based on upper record statistics values. Bayes estimations of SEL and linear exponential loss (LINEX) functions and maximum likelihood estimation (MLE) are derived for linear exponential distribution based on SRS and record ranked set sampling (RRSS). These estimators are compared via their bias and mean squared error (MSE). A simulation study and real data are carried out to study the precision of MLE and Bayesian estimations for the parameters involved. The results of this chapter were accepted for publication at:

"Journal of Mathematics and Statistics, to appear."

Chapter 5

In this chapter, we use ORSS from order statistics of INID random variables to obtain Bayesian estimation for the scale parameter of Rayleigh distribution under Type-II doubly censoring scheme. This is done with respect to both SEL and Al-Bayyati loss functions (ALF). We obtain MSE and bias of the derived estimators based on ORSS and compare them with those based on the corresponding SRS to appreciate the efficiency of the obtained estimators. Furthermore, we present the two-sample Bayesian predictive density function (point and interval) for the ordered future sample. Finally, a simulation study and real data are conducted to assess the performance of the theoretical results. The results of this chapter were accepted for publication at:

"American Journal of Statistics and Probability, to appear."

Chapter 6

In this chapter, order statistics of INID random variables are used to obtain ORSS. Recurrence relations for single and product moments are obtained from Kumaraswamy generalized distribution based on ordinary order statistics. The Kumaraswamy generalized distribution has a large number of well known lifetime special submodels such as the Kumaraswamy Weibull, Kumaraswamy log-generalized inverse Weibull and Kumaraswamy generalized Rayleigh distributions, among others. The Kumaraswamy log-generalized inverse Weibull distribution is given as an application to obtain best linear unbiased estimators (BLUEs) of the location and scale parameters using ORSS and SRS. The relative efficiency of the derived estimators are obtained to compare (BLUEs) based on ORSS (BLUEs-ORSS) with BLUEs-SRS. We show that BLUEs-ORSS are better than BLUEs-SRS for the location and scale parameters of Kumaraswamy log-generalized inverse Weibull distribution. The results of this chapter were submitted to an international statistical journal.

Chapter 1

Introduction

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

1.1 Order statistics

Consider $Y_1, Y_2, \ldots, Y_n$ be n independent identically distributed (iid) random variables with cdf $F(y)$ and pdf $f(y)$ then if we arrange these in ascending order of magnitude such as $y_{1:n} \leq y_{2:n} \leq \ldots \leq y_{n:n}$, then $Y_{j:n}$ $(j = 1, 2, \ldots, n)$ is called the jth order statistic (OS) from a random sample of size n. The pdf and cdf of jth order statistic, $1 \leq j \leq n$ can be written respectively, as

$$f_{j:n}(y) = \frac{n!}{(j-1)!(n-j)!}(F(y))^{j-1}(\overline{F}(y))^{n-j}f(y), \quad -\infty < y < \infty, \tag{1.1}$$

$$F_{j:n}(y) = \sum_{i=j}^{n} \binom{n}{i} (F(y))^{i}(1-F(y))^{n-i}, \tag{1.2}$$

and

$$\overline{F}_{j:n}(y) = \sum_{i=0}^{r-1} \binom{n}{i} (F(y))^{i}(1-F(y))^{n-i}. \tag{1.3}$$

We can re-write the pdf and cdf in Eqs. (1.1) and (1.2) respectively, using the binomial expansion in the following form

$$f_{j:n}(y) = \sum_{i=0}^{j-1} \widetilde{c}_{i,j}(n)(1-F(y))^{n+i-j} f(y), \tag{1.4}$$

and

$$F_{j:n}(y) = \sum_{i=j}^{n} \sum_{\tau=0}^{i} c_{i,\tau}(n)(1-F(y))^{n+\tau-i}, \tag{1.5}$$

or, equivalently,

$$F_{j:n}(y) = 1 - \sum_{i=1}^{j} \omega_{i,j}(n)\left(1-F(y)\right)^{n+i-j}, \tag{1.6}$$

where $\widetilde{c}_{i,j}(n) = (-1)^i j \binom{j-1}{i} \binom{n}{j}$, $c_{i,\tau}(n) = \binom{n}{i}\binom{i}{\tau}(-1)^\tau$ and $\omega_{i,j}(n) = \widetilde{c}_{i-1,j}(n)/(n+i-j)$.

The joint pdf of $Y_{r:n}$ and $Y_{s:n}$, $1 \leq r < s \leq n$ is given by

$$\begin{aligned} f_{r,s:n}(x,y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left(F(x)\right)^{r-1} f(x) \\ &\times \left(F(y)-F(x)\right)^{s-r-1}\left(1-F(y)\right)^{n-s} f(y), \quad -\infty < x < y < \infty. \end{aligned} \tag{1.7}$$

Also, the joint pdf of $Y_{1:n}, Y_{2:n}, ..., Y_{n:n}$ is given by

$$f_{1,2,...,n:n}(y_1, y_2, ..., y_n) = n! \prod_{r=1}^{n} f(y_r), \quad -\infty < y_1 \leq y_2 \leq ... \leq y_n < \infty. \tag{1.8}$$

For more details about order statistics, see Arnold et al. [13]; David and Nagaraja [27].

1.1.1 Moments of order statistics

From Eq. (1.1), the ℓth moments of order statistic denoted by $\mu_{r:n}^{(\ell)}$, $\ell = 1, 2, ..., 1 \le r \le n$ is given by

$$\begin{aligned} \mu_{r:n}^{(\ell)} &= E[X_{r:n}^{\ell}] \\ &= \frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{\infty} x^{\ell} (F(x))^{r-1} (\overline{F}(x))^{n-r} \\ &\times \ f(x) \mathrm{d}x, \quad -\infty < x < \infty. \end{aligned} \tag{1.9}$$

Also, by using Eq. (1.7), the (ℓ, b) product moment of $X_{r:n}$ and $X_{s:n}$ for ℓ, b are positive integers is given by

$$\begin{aligned} \mu_{r,s:n}^{(\ell,b)} &= E[X_{r:n}^{\ell} X_{s:n}^{b}] \\ &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \int_{-\infty}^{\infty} \int_{-\infty}^{y} x^{\ell} y^{b} \left(F(x)\right)^{r-1} f(x) \\ &\times \ \left(F(y) - F(x)\right)^{s-r-1} \left(1 - F(y)\right)^{n-s} f(y) \mathrm{d}x \mathrm{d}y, \end{aligned} \tag{1.10}$$

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

1.2 Record values

Record values are used in many parts of statistics such as industrial stress testing, sporting and athletic events, oil and mining surveys. Suppose that $X_1, X_2, \ldots$ be an infinite iid random variables having cdf $F(x)$ and corresponding pdf $f(x)$, so the observation X_j is called an upper RVs if its value is more than of all previous observations and is called an lower RVs if its value is less than of all previous observations.

1.2.1 Ordinary record values

Let $\{X_i, i \ge 1\}$ be sequences of iid random variables having cdf $F(x)$ and corresponding pdf $f(x)$ and the sequence $\{Y_i, i \ge 1, k \ge 1\}$ be the upper record values of $\{X_i, i \ge 1\}$. So the joint density function of Y_i, $i = 1, 2, ..., n$ is in the form

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = f(y_n) \prod_{i=1}^{n-1} \frac{f(y_i)}{\overline{F}(y_i)}. \tag{1.11}$$

The pdf of Y_r is given by

$$f_{Y_r}(y) = \frac{1}{(r-1)!}(-\log(\overline{F}(y)))^{r-1} f(y). \tag{1.12}$$

The joint pdf of Y_r and Y_n is given by

$$\begin{aligned} f_{Y_r,Y_n}(x,y) &= \frac{1}{(r-1)!(n-r-1)!}(-\log(\overline{F}(y)) + \log(\overline{F}(x)))^{n-r-1} \\ &\times (-\log(\overline{F}(x)))^{r-1} \frac{f(x)}{\overline{F}(x)} f(y). \end{aligned} \tag{1.13}$$

For more details about records, see Arnold et al. [14].

1.3 Lifetime data

The lifetime of an item or system is the period of time during which item or system functions. The lifetime data observed from a life testing experiment could be separated into two categories: complete (all failure data are available) or censored (some of failure data are not observed in the study time). The second type is common use in life testing and reliability experiments to decrease the lifetime of testing and reduce the test expense.

1.3.1 Complete data

Complete data is occured when the failure time of each sample unit is observed or known. The failure data observed from a life testing arise in a naturally increasing order. Therefore, order statistics is used to analyze types of lifetime data. Let $x_1, x_2, \cdots, x_n$ be a random sample of size n from cdf $F(x)$ and pdf $f(x)$. In this case we have $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, $x_1 \leq x_2 \leq \cdots \leq x_n$. So the joint density of $\mathbf{x}$ is written as

$$f(\mathbf{x}) = n! \prod_{i=1}^{n} f(x_i), \quad 0 < x_1 \leq x_2 \leq \cdots \leq x_n < \infty. \tag{1.14}$$

1.3.2 Censored data

Suppose that n items or equipments are put on a life test, the experiment is terminated when only r out of n are failed. In general, the failure data are