

Order Statistics and Inference



Order Statistics and Inference

Estimation Methods

N. Balakrishnan

*Department of Mathematics and Statistics
McMaster University
Hamilton, Ontario
Canada*

and

A. Clifford Cohen

*Department of Statistics
University of Georgia
Athens, Georgia*



ACADEMIC PRESS, INC.

Harcourt Brace Jovanovich, Publishers

Boston San Diego New York

London Sydney Tokyo Toronto

This book is printed on acid-free paper. ∞

Copyright © 1991 by Academic Press, Inc.

All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

ACADEMIC PRESS, INC.

1250 Sixth Avenue, San Diego, CA 92101

United Kingdom Edition published by

ACADEMIC PRESS LIMITED

24-28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data

Balakrishnan, N., date.

Order statistics and inference: estimation methods/N.

Balakrishnan, A. Clifford Cohen.

p. cm. — (Statistical modeling and decision science)

Includes bibliographical references (p. ??) and index.

ISBN 0-12-076948-4 (alk. paper)

1. Order statistics. 2. Estimation theory.

I. Cohen, A. Clifford. II. Title. III. Series

QA278.7.B35 1990

519.2—dc20

90-824

CIP

90 91 92 93 9 8 7 6 5 4 3 2 1

Printed in the United States of America

Acknowledgments

The able and generous assistance of Edna Pathmanathan (McMaster University) and Gayle Rodriguez, Molly Rema, and Dawn Tolbert (University of Georgia) in typing the manuscript is acknowledged with thanks. Special thanks are extended to Mrs. Susan Gay (Assistant Editor, Academic Press, Boston) and Ms. Amy Strong (Production Editor) for their kind cooperation and prompt responses during the preparation of this book.

Thanks are offered to the American Statistical Association, Marcel Dekker, Inc., the American Society for Quality Control, Gordon and Breach Science Publishers, Inc., the Institute of Electrical and Electronics Engineers, Inc., the Biometrika Trustees, and the Editor of the *Journal of Statistical Planning and Inference*, for permission to reproduce previously published tables, charts, and examples.

The first author thanks the Natural Sciences and Engineering Research Council of Canada and the Science and Engineering Research Board of McMaster University for providing partial financial support (toward the typing and the computational costs) during the preparation of this book.

The second author extends thanks to Dr. Lynne Billard (former Head) and Dr. Robert L. Taylor (present Head) for encouragement and for making facilities of the Department of Statistics at the University of Georgia available during the preparation of this book, and to his colleague and former student, Dr. Betty Jones Whitten, for her assistance in computing, proofreading, and editing, and also for the encouragement she has given during the course of this project.

Preface

The subject of order statistics has been in the process of development for many years and recently has become increasingly important. Articles relating to this area have appeared in numerous different publications. In this volume, we have made a sincere effort to consolidate important developments in methods of estimation and to illustrate them with practical examples from various scientific disciplines. In addition to an account of classical moment and maximum likelihood estimation, emphasis has been placed on linear unbiased and optimal linear estimation based on complete as well as censored samples. Recent developments in simple approximate maximum likelihood estimation based on censored samples as well as the sample completion technique for censored data have been included and explained through several examples.

This volume is intended both as a text and as a reference source for researchers and practitioners in order statistics, estimation theory, life testing, quality control, and reliability. For this purpose, an up-to-date, comprehensive bibliography on this topic has been included. We have also incorporated numerous illustrative examples which will be of particular benefit to students and practitioners.

In reporting on various developments in this area, we have tried to the best of our knowledge to give credit and to recognize priority where appropriate for previously published results. We apologize for any oversights

and will cheerfully accept corrections. We will be grateful for any information regarding omission of relevant references.

N. Balakrishnan
Hamilton, Ontario
Canada

A. Clifford Cohen
Athens, Georgia
U.S.A.

March 1990

Contents

<i>Acknowledgments</i>	xi
<i>Preface</i>	xiii
<i>List of Tables</i>	xv
<i>List of Figures</i>	xix
 Chapter 1 Introduction	 1
1.1 Introductory Remarks	1
1.2 The Role of Order Statistics in Practical Applications	2
1.3 Scope of This Volume	4
 Chapter 2 Basic Theory	 7
2.1 Introduction	7
2.2 Joint Distribution of n Order Statistics	7
2.3 Joint Distribution of Two Order Statistics	8
2.4 Distribution of a Single Order Statistic	11
2.5 Distribution of Range and Some Other Statistics	17
 Chapter 3 Moments and Other Expected Values	 21
3.1 Introduction	21
3.2 Some Basic Formulas	22
3.3 Recurrence Relations and Identities	23
3.4 Results for the Uniform Distribution	30
3.5 Results for the Exponential Distribution	34
	vii

3.6	Results for the Logistic Distribution	38
3.7	Results for the Gamma Distribution	45
3.8	Results for the Weibull Distribution	47
3.9	Results for the Normal Distribution	51
3.10	Results for the Half Logistic Distribution	63
3.11	David and Johnson's Approximation	68

Chapter 4 Linear Estimation Based on Order Statistics 73

4.1	Preliminary Remarks	73
4.2	BLUE of the Scale Parameter	74
4.3	BLUE for the One-Parameter Exponential Distribution	76
4.4	BLUEs of the Location and Scale Parameters	80
4.5	BLUEs for the Two-Parameter Exponential Distribution	87
4.6	Gupta's Simplified Linear Estimators	94
4.7	Blom's Unbiased Nearly Best Linear Estimators	100
4.8	Downton's Linear Estimators with Polynomial Coefficients ...	109
4.9	Details of Other Related Work	119

Chapter 5 Maximum Likelihood Estimation 121

5.1	Preliminary Remarks	121
5.2	The Weibull Distribution	122
5.3	The Lognormal Distribution	126
5.4	The Inverse Gaussian Distribution	131
5.5	The Gamma Distribution	135
5.6	The Rayleigh Distribution	139
5.7	The Exponential Distribution	144
5.8	Complete, Truncated, and Censored Samples from the Normal Distribution	146

Chapter 6 Approximate Maximum Likelihood Estimation 161

6.1	Preliminary Remarks	161
6.2	Estimation for the Rayleigh Distribution	162
6.3	Estimation for the Normal Distribution	167
6.4	Estimation for the Logistic Distribution	177
6.5	Estimation for the Extreme Value Distribution	186
6.6	Estimation for the Type I Generalized Logistic Distribution ...	197
6.7	Estimation for the Half Logistic Distribution	208

Chapter 7 Optimal Linear Estimation Based on Selected Order Statistics	215
7.1 Introduction	215
7.2 Bennett's and Jung's Optimal Asymptotic Estimators	216
7.3 Ogawa's Optimal Estimators Based on Selected Order Statistics	233
7.4 Dixon's Simplified Linear Estimators	249
7.5 Balakrishnan's Approximate Maximum Likelihood Estimation	255
7.6 Estimation of Population Quantiles	264
7.7 Details of Other Related Work	270
Chapter 8 Cohen-Whitten Estimators: Using Order Statistics	273
8.1 Preliminary Remarks	273
8.2 The Weibull Distribution	274
8.3 The Lognormal Distribution	278
8.4 The Inverse Gaussian Distribution	281
8.5 The Gamma Distribution	286
8.6 The Exponential Distribution	289
8.7 Illustrative Examples	290
Chapter 9 Estimation in Regression Models	299
9.1 Introduction	299
9.2 Best Linear Unbiased Estimation with Multiple Measurements	300
9.3 Modified Maximum Likelihood Estimation with Multiple Measurements	310
9.4 Modified Maximum Likelihood Estimation with Single Measurements	315
Chapter 10 A Sample Completion Technique for Censored Samples	329
10.1 Introductory Remarks	329
10.2 Censored Samples from the Normal Distribution	330
10.3 Censored Samples from Skewed Distributions	331
10.4 Illustrative Examples	335
<i>Bibliography</i>	341
<i>Author Index</i>	367
<i>Subject Index</i>	373

List of Tables

4.4.1	The BLUEs of μ and σ and Their Standard Errors for the Data of D'agostino and Stephens	85
4.5.1	Days of Incubation among 10 Rabbits Following Inoculation with Graded Amounts of <i>Treponema pallidum</i> ; BLUEs	93
4.6.1	Comparison of the Estimators $\tilde{\sigma}$, σ^* , and σ^{**} for the Normal Population	97
4.8.1	Coefficients $b_{ij}^{(s)}$ for Computing Elements of Ω in the Normal Distribution	113
4.8.2	Efficiency of Range as an Estimate of σ in the Normal Distribution relative to $\sigma_{**} = C \sum_{i=1}^n (2i - n - 1) X_{i:n} / \{n(n-1)\} [C = 1.77245385] \dots\dots\dots$	114
4.8.3	Coefficients $b_{ij}^{(s)}$ for Computing Elements of Ω in the Logistic Distribution	116
4.8.4	Values of Ω_{ij} for the Logistic Distribution When $n = 20$	116
4.8.5	Coefficients, Variances, and Relative Efficiencies of the Estimators μ_{**} and σ_{**} When $n = 20$	117
5.2.1	Variance-Covariance Factors for Maximum Likelihood Estimates of Weibull Parameters	125
5.3.1	Variance-Covariance Factors for Maximum Likelihood Estimates of Lognormal Parameters	130
5.4.1	Variance-Covariance Factors for Maximum Likelihood Estimates of Inverse Gaussian Parameters	134
5.5.1	Variance-Covariance Factors for Maximum Likelihood Estimates of Gamma Distribution Parameters	138

5.8.1	Auxiliary Estimation Function θ for Singly Truncated Normal Samples	150
5.8.2	Auxiliary Estimation Function $\lambda(h, \alpha)$ for Singly Censored Normal Samples	154
5.8.3	Variance Factors for Singly Truncated and Singly Censored Normal Samples	156
6.2.1	Bias, Variance, and Approximate Variance of $\hat{\sigma}$, and Variance of the BLUE of σ for the Rayleigh Distribution	166
6.3.1	Bias, Variances, and Covariance of Estimators $\hat{\mu}$ and $\hat{\sigma}$ for the Normal Distribution	172
6.3.2	Simulated Values of Bias, Variances, and Covariance of $\hat{\mu}$ and $\hat{\sigma}$ for the Normal Distribution	173
6.3.3	Variances of the Estimators of μ for the Normal Distribution When $r = s$	174
6.4.1	Bias, Variances, and Covariance of Estimators $\hat{\mu}$ and $\hat{\sigma}$ for the Logistic Distribution	182
6.4.2	Simulated Values of Bias, Variances, and Covariance of $\hat{\mu}$ and $\hat{\sigma}$ for the Logistic Distribution	183
6.4.3	Variances of the Estimators of μ for the Logistic Distribution When $r = s$	183
6.5.1	Values of Bias, Variances, and Covariance of $\hat{\mu}$ and $\hat{\sigma}$ for the Extreme Value Distribution When $n = 10$	192
6.5.2	Values of Bias, Variances, and Covariance of $\hat{\mu}$ and $\hat{\sigma}$ for the Extreme Value Distribution When $n = 20$	193
6.5.3	Comparison of the Bias and Mean Square Error of Various Estimators of μ and σ for the Extreme Value Distribution	194
6.6.1	Bias, Variances, and Covariance of $\hat{\mu}$ and $\hat{\sigma}$ for the Generalized Logistic Distribution When $n = 15$ and $b = 1.0$	203
6.6.2	Bias, Variances, and Covariance of $\hat{\mu}$ and $\hat{\sigma}$ for the Generalized Logistic Distribution When $n = 15$ and $b = 2.0$	204
6.6.3	Bias, Variances, and Covariance of $\hat{\mu}$ and $\hat{\sigma}$ for the Generalized Logistic Distribution When $n = 15$ and $b = 3.0$	205
6.6.4	Comparison of $\hat{\mu}$ and $\hat{\sigma}$ with BLUEs μ^* and σ^* for the Generalized Logistic Distribution	206
6.6.5	Values of $\hat{\mu}$ and $\hat{\sigma}$ and Their Standard Errors for Censored Samples from the Generalised Logistic Distribution	207
6.7.1	Unbiasing Factor of the AMLE and Variance of the Unbiased AMLE, Unbiasing Factor of the MLE and Variance of the Unbiased MLE, and Variance of the BLUE of σ for the Half Logistic Distribution When $n = 10$	210
6.7.2	Unbiasing Factor of the AMLE and Variance of the Unbiased AMLE, Unbiasing Factor of the MLE and Variance of the Unbiased MLE, and Variance of the BLUE of σ for the Half Logistic Distribution When $n = 15$	211

6.7.3	Unbiasing Factor of the AMLE and Variance of the Unbiased AMLE, Unbiasing Factor of the MLE and Variance of the Unbiased MLE, and Variance of the BLUE of σ for the Half Logistic Distribution When $n = 20$	212
7.3.1	Optimum Spacings, the Corresponding Coefficients, and Relative Efficiencies for $\hat{\sigma}^*$ of the One-Parameter Exponential Distribution	241
7.3.2.	Time Intervals in Days between Explosions in Mines Involving More Than 10 Men Killed from 6 December 1875 to 29 May 1951	250
7.4.1	Estimator in (7.4.1) for the Mean of the Normal Population and Values of Variance and Efficiency Relative to the Sample Mean	251
7.4.2	Estimator in (7.4.3) for the Standard Deviation of the Normal Population and Values of Variance and Relative Efficiencies	252
7.4.3	Estimator in (7.4.4) for the Standard Deviation of the Normal Population and Values of Variance and Relative Efficiencies	253
7.4.4	Estimator in (7.4.1) for the Mean of the Logistic Population and Values of Variance and Efficiency Relative to the BLUE	254
7.4.5	Estimator in (7.4.3) for the Standard Deviation of the Logistic Population and Values of Variance and Efficiency Relative to the BLUE ..	255
7.4.6	Estimator in (7.4.1) for the Mean of the Double Exponential Population and Values of Variance and Efficiency Relative to the BLUE ..	256
7.4.7	Estimator in (7.4.3) for σ of the Double Exponential Population and Values of Variance and Efficiency Relative to the BLUE	257
8.3.1	Expected Values of the First Order Statistics from the Standard Normal Distribution (0, 1)	281
8.7.1	Maximum Flood Levels of Susquehanna River at Harrisburg, PA over Four-Year periods (1890-1969) in Millions of Cubic Feet per Second. Example 8.7.1	291
8.7.2	Fatigue Life in Hours of a Certain Type of Bearing. Example 8.7.2	291
8.7.3	Random Sample from an Inverse Gaussian Distribution: $E(X) = 16$, $V(X) = 9$, $\alpha_3(X) = 1.5$. Example 8.7.3	292
8.7.4	Summary of Estimates for Example 8.7.1. $\bar{x} = 0.423125$, $s = 0.1252789$, $a_3 = 1.0673243$, $x_1 = 0.265$, $n = 20$, $s^2/(\bar{x} - x_1)^2 = 0.6277$, $z_1 = -1.2622$	293
8.7.5	Summary of Estimates for Example 8.7.2. $\bar{x} = 220.48$, $s = 78.405638$, $a_3 = 1.8635835$, $x_1 = 152.7$, $n = 10$, $s^2/(\bar{x} - x_1)^2 = 1.338$, $z_1 = -0.8645$	294
8.7.6	Summary of Estimates for Example 8.7.3. $\bar{x} = 15.7415$, $s = 3.1276$, $a_3 = 1.3230$, $x_1 = 11.3557$, $n = 100$, $s^2/(\bar{x} - x_1)^2 = 0.50854$, $z_1 = -1.4023$	294
9.2.1	Sales (y) by Months on the Job (x)	305
9.2.2	BLUES of α , β , and σ and Their Variances for Different Distributions	306
9.3.1	The Log Survival Times of the Specimens at Five Doses	313
9.3.2	Computational Tableau for the Evaluation of the MMLEs	313
9.4.1	Values of Bias and Variance of the MLEs and the MMLEs of α , β , and σ ; $n = 10$	323

9.4.2	Values of Bias and Variance of the MLEs and the MMLEs of α , β , and σ ; $n = 20$	324
9.4.3	Normalized Residuals and Their Ranks for Daniel and Wood's Data	325
9.4.4	Normalized Residuals and Their Ranks for Daniel and Wood's Data with Contamination	327
9.4.5	Mendenhall's EPA 1980 Mileage Rating Data	328
10.4.1	Random Sample from a Normal Distribution, $N = 100$, $\mu = 60$, and $\sigma = 10$	336
10.4.2	Censored Sample Estimates of Normal Distribution Parameters	337
10.4.3	Successive Iterations of Pseudo-Complete Sample Estimates of Normal Distribution Parameters; $N = 100$, $n = 95$, $c = 5$, $h = 0.05$	337
10.4.4	Random Sample from an Inverse Gaussian Distribution; $\gamma = 10$, $\mu = 6$, $\sigma = 5$ ($\alpha_3 = 2.5$)	338
10.4.5	Summary of Estimates of Inverse Gaussian Distribution Parameters	338
10.4.6	Successive Iterations of Pseudo-Complete Sample Estimates of Inverse Gaussian Parameters; $\gamma = 10$, $\mu = 6$, $\sigma = 5$, $E(X) = 16$, $\alpha_3 = 2.5$, $N = 100$, $h = 0.05$	339

List of Figures

5.6.1	Truncated sample estimating function for Rayleigh distribution	142
5.8.1	Estimation curve for singly truncated normal samples	149
8.2.1	Graphs of $C(\delta)$ and $D(\delta)$ for the Weibull distribution	276
8.2.2	Graphs of $W(n, \hat{\delta}) = s^2/(\bar{x} - x_1)^2$ for the Weibull distribution	277
8.3.1	Graphs of $\omega(\omega - 1)/[\sqrt{\omega} - \exp\{\sigma E(Z_{1:n})\}]^2$ for the lognormal distribution	282
8.4.1	Graphs of α_3 as a function of z_1 and n for the inverse Gaussian distribution	285
8.5.1	Graphs of α_3 as a function of z_1 and n for the gamma distribution	288

Chapter 1

Introduction

1.1. Introductory Remarks

The subject of order statistics deals with properties and applications of ordered random variables and of functions of these variables. If the random variables $\{X_i\}$, $i = 1, 2, \dots, n$ are arranged in ascending order of magnitude and then written as

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

then $X_{i:n}$ is said to be the i th-order statistic in a sample of size n . In the usual random sampling theory, the unordered X_i are assumed to be statistically independent and identically distributed. Because of the inequality relations among them, the order statistics $X_{i:n}$ are necessarily dependent. A distinction is made between the random variables $X_{i:n}$ and the corresponding sample observations $x_{i:n}$. As an illustration, consider a life test on a certain electrical component. A random selection of n specimens is made from the population of items at issue. This sample is placed "on test." If the test is continued until all sample specimens have failed, the sample is said to be complete, and it consists of the naturally ordered observations

$$x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}.$$

If sample specimens are withdrawn prior to failure and/or testing is terminated with survivors, the sample is said to be censored. In some situations, the original sample observations x_i may be unordered with respect to

magnitude, and a transformation — that is, a rearrangement — is required to produce a corresponding ordered sample.

Some frequently encountered functions of order statistics are the *extremes* $X_{1:n}$ and $X_{n:n}$, the *range* $W = X_{n:n} - X_{1:n}$, the *extreme deviate* from the sample mean, $X_{n:n} - \bar{X}$, and for a random sample from a normal distribution $N(\mu, \sigma^2)$, the *studentized range*, W/S_ν , where S_ν is a root-mean-square estimator of σ based on ν degrees of freedom. All of these statistics have important applications. The extremes arise in the statistical study of floods and droughts, as well as in breaking strength and fatigue failure studies. The range is widely employed in the field of quality control as a quick estimator of the standard deviation σ . The extreme deviate is a basic tool in procedures for detecting outliers. Large values of $(X_{n:n} - \bar{X})/\sigma$ suggest the presence of outliers. When outliers are not confined to one direction, the studentized range is also useful in the detection process. Furthermore, the studentized range is the basis of many quick tests in small samples, and it is of key importance in ranking "treatment" means in analysis of variance situations.

With regard to notation, random variables are designated by uppercase letters, and their realizations (observations) by corresponding lowercase letters, as shown below.

X_1, X_2, \dots, X_n	unordered random variables
x_1, x_2, \dots, x_n	unordered observations
$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$	ordered random variables
$x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$	ordered observations.

By order statistics, we mean either ordered random variables or ordered observations. In some applications where only the first order statistic is involved, the notation will be abbreviated from $X_{1:n}$ to simply X_1 , and from $x_{1:n}$ to x_1 . This shortened notation will be employed only where no confusion is likely to result.

1.2. The Role of Order Statistics in Practical Applications

Order statistics and functions of these statistics play an important role in numerous practical applications. As previously mentioned, the range is

widely used by quality control practitioners to provide quick estimates of σ in the normal distribution. The first- and the n th-order statistics are valuable tools for the detection of outliers. The studentized range is also useful in the detection of outliers. However, its principal use is as the basis of quick tests in small samples and in ranking procedures.

Sarhan and Greenberg (1962) employed linear functions of order statistics in conjunction with the Gauss-Markov theorem to systematically estimate location and scale parameters in both complete and censored samples. They provided tables of the coefficients necessary for the calculations of these estimates from samples varying in size from 2 to 20.

Other applications of order statistics arise in the study of system reliability. A system of n components is called a k -out-of- n system if it remains operational only if at least k components continue to function. For components with independent lifetime distributions, the time to failure of the system is thus the $(n - k + 1)$ th-order statistic. The special cases $k = 1$ and $k = n$ correspond respectively to parallel and series systems.

A major impetus for the study of order statistics has been provided by the development of modern computers. Through their use it is feasible to make repeated examinations of the same data in many different ways. Tukey (1970), Mosteller and Tukey (1977), and others have employed various informal techniques in the analysis of data. It is thus possible to determine quickly if the data indeed are in accord with an assumed distribution and with an assumed model. A plot of the ordered observations against some simple function of their ranks, preferably on probability paper appropriate for the assumed distribution, will often prove helpful in making such determinations. Nelson (1972) has exploited this technique to obtain parameter estimates from progressively censored samples by plotting hazard functions. A straight-line fit to either a probability or a hazard function plot indicates that both the model and the parameter estimates are generally satisfactory, whereas serious departures from a straight line reveal either the presence of outliers or the selection of an incorrect model.

The first-order statistic is particularly useful in estimating the threshold parameter in skewed distributions. Cohen (1988), Cohen and Whitten (1980, 1981, 1982, 1985, 1986, 1988), and Cohen *et al.* (1984, 1985) have employed the first-order statistic to advantage in various modified moment and maximum likelihood estimators of parameters in the Weibull, lognormal, inverse Gaussian, gamma, generalized gamma, and other positively skewed distributions.

A somewhat more limited application of order statistics can be found in the area of data compression, in which large masses of data are replaced