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

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 90-824
519.2—dc20 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

List of Tables

List of Figures

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 $\{X_i\}$, $i = 1, 2, \ldots, n$ are arranged in ascending order of magnitude and then written as

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

then $X_{i:n}$ is said to be the ith-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

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

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.

$X_1, X_2, \ldots, X_n$	unordered random variables
$x_1, x_2, \ldots, x_n$	unordered observations
$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$	ordered random variables
$x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$	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 nth-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