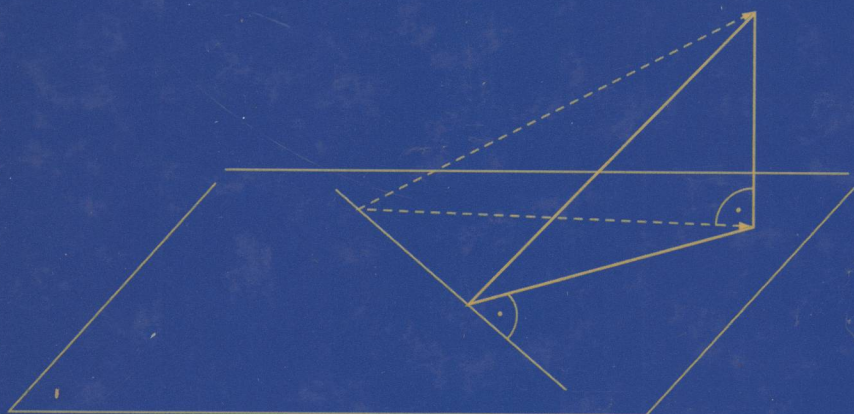


SPRINGER TEXTS IN STATISTICS

Testing Statistical Hypotheses

THIRD EDITION



E.L. Lehmann
Joseph P. Romano



Springer

0212.1
L523
E.3
E.L. Lehmann Joseph P. Romano

Testing Statistical Hypotheses

Third Edition



 Springer



E.L. Lehmann
Professor of Statistics Emeritus
Department of Statistics
University of California, Berkeley
Berkeley, CA 94720
USA

Joseph P. Romano
Department of Statistics
Stanford University
Sequoia Hall
Stanford, CA 94305
USA
romano@stanford.edu

Editorial Board

George Casella
Department of Statistics
University of Florida
Gainesville, FL 32611-8545
USA

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress.

ISBN 0-387-98864-5

Printed on acid-free paper.

© 2005, 1986, 1959 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3

springer.com

Springer Texts in Statistics

Advisors:

George Casella Stephen Fienberg Ingram Olkin

Springer Texts in Statistics

- Alfred*: Elements of Statistics for the Life and Social Sciences
Berger: An Introduction to Probability and Stochastic Processes, Second Edition
Bilodeau and Brenner: Theory of Multivariate Statistics
Blom: Probability and Statistics: Theory and Applications
Brockwell and Davis: Introduction to Times Series and Forecasting, Second Edition
Carmona: Statistical Analysis of Financial Data in S-Plus
Chow and Teicher: Probability Theory: Independence, Interchangeability, Martingales, Third Edition
Christensen: Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization, Second Edition
Christensen: Log-Linear Models and Logistic Regression, Second Edition
Christensen: Plane Answers to Complex Questions: The Theory of Linear Models, Third Edition
Creighton: A First Course in Probability Models and Statistical Inference
Davis: Statistical Methods for the Analysis of Repeated Measurements
Dean and Voss: Design and Analysis of Experiments
du Toit, Steyn, and Stumpf: Graphical Exploratory Data Analysis
Durrett: Essentials of Stochastic Processes
Edwards: Introduction to Graphical Modelling, Second Edition
Everitt: An R and S-Plus® Companion to Multivariate Analysis
Finkelstein and Levin: Statistics for Lawyers
Flury: A First Course in Multivariate Statistics
Gut: Probability: A Graduate Course
Heiberger and Holland: Statistical Analysis and Data Display: An Intermediate Course with Examples in S-PLUS, R, and SAS
Jobson: Applied Multivariate Data Analysis, Volume I: Regression and Experimental Design
Jobson: Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods
Kalbfleisch: Probability and Statistical Inference, Volume I: Probability, Second Edition
Kalbfleisch: Probability and Statistical Inference, Volume II: Statistical Inference, Second Edition
Karr: Probability
Keyfitz: Applied Mathematical Demography, Second Edition
Kiefer: Introduction to Statistical Inference
Kokoska and Nevison: Statistical Tables and Formulae
Kulkarni: Modeling, Analysis, Design, and Control of Stochastic Systems
Lange: Applied Probability
Lange: Optimization
Lehmann: Elements of Large-Sample Theory

(continued after index)

Dedicated to the Memory of

Lucien Le Cam (1924-2000) and John W. Tukey (1915-2000)

Preface to the Third Edition

The Third Edition of *Testing Statistical Hypotheses* brings it into consonance with the Second Edition of its companion volume on point estimation (Lehmann and Casella, 1998) to which we shall refer as *TPE2*. We won't here comment on the long history of the book which is recounted in Lehmann (1997) but shall use this Preface to indicate the principal changes from the 2nd Edition.

The present volume is divided into two parts. Part I (Chapters 1–10) treats small-sample theory, while Part II (Chapters 11–15) treats large-sample theory. The preface to the 2nd Edition stated that “the most important omission is an adequate treatment of optimality paralleling that given for estimation in *TPE*.” We shall here remedy this failure by treating the difficult topic of asymptotic optimality (in Chapter 13) together with the large-sample tools needed for this purpose (in Chapters 11 and 12). Having developed these tools, we use them in Chapter 14 to give a much fuller treatment of tests of goodness of fit than was possible in the 2nd Edition, and in Chapter 15 to provide an introduction to the bootstrap and related techniques. Various large-sample considerations that in the Second Edition were discussed in earlier chapters now have been moved to Chapter 11.

Another major addition is a more comprehensive treatment of multiple testing including some recent optimality results. This topic is now presented in Chapter 9. In order to make room for these extensive additions, we had to eliminate some material found in the Second Edition, primarily the coverage of the multivariate linear hypothesis.

Except for some of the basic results from Part I, a detailed knowledge of small-sample theory is not required for Part II. In particular, the necessary background should include: Chapter 3, Sections 3.1–3.5, 3.8–3.9; Chapter 4: Sections 4.1–4.4; Chapter 5, Sections 5.1–5.3; Chapter 6, Sections 6.1–6.2; Chapter 7, Sections 7.1–7.2; Chapter 8, Sections 8.1–8.2, 8.4–8.5.

Of the two principal additions to the Third Edition, multiple comparisons and asymptotic optimality, each has a godfather. The development of multiple comparisons owes much to the 1953 volume on the subject by John Tukey, a mimeographed version which was widely distributed at the time. It was officially published only in 1994 as Volume VIII in *The Collected Works of John W. Tukey*.

Many of the basic ideas on asymptotic optimality are due to the work of Le Cam between 1955 and 1980. It culminated in his 1986 book, *Asymptotic Methods in Statistical Decision Theory*.

The work of these two authors, both of whom died in 2000, spans the achievements of statistics in the second half of the 20th century, from model-free data analysis to the most abstract and mathematical asymptotic theory. In acknowledgment of their great accomplishments, this volume is dedicated to their memory.

Special thanks to George Chang, Noureddine El Karoui, Matt Finkelman, Brit Katzen, Mee Young Park, Elizabeth Purdom, Armin Schwartzman, Azeem Shaikh and the many students at Stanford University who proofread several versions of the new chapters and worked through many of the over 300 new problems. The support and suggestions of our colleagues is greatly appreciated, especially Persi Diaconis, Brad Efron, Susan Holmes, Balasubramanian Narasimhan, Dimitris Politis, Julie Shaffer, Guenther Walther and Michael Wolf. Finally, heartfelt thanks go to friends and family who provided continual encouragement, especially Joe Chavez, Ann Marie and Mark Hodges, David Fogle, Scott Madover, Tom Neville, David Olachea, Janis and Jon Squire, Lucy, and Ron Susek.

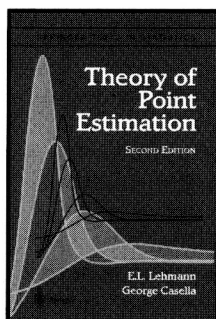
E. L. Lehmann
Joseph P. Romano

January, 2005

Springer Texts in Statistics (continued from page ii)

- Lehmann*: Testing Statistical Hypotheses, Second Edition
Lehmann and Casella: Theory of Point Estimation, Second Edition
Lindman: Analysis of Variance in Experimental Design
Lindsey: Applying Generalized Linear Models
Madansky: Prescriptions for Working Statisticians
McPherson: Applying and Interpreting Statistics: A Comprehensive Guide, Second Edition
Mueller: Basic Principles of Structural Equation Modeling: An Introduction to LISREL and EQS
Nguyen and Rogers: Fundamentals of Mathematical Statistics: Volume I: Probability for Statisticians
Nguyen and Rogers: Fundamentals of Mathematical Statistics: Volume II: Statistical Inference
Noether: Introduction to Statistics: The Nonparametric Way
Nolan and Speed: Stat Labs: Mathematical Statistics Through Applications
Peters: Counting for Something: Statistical Principles and Personalities
Pfeiffer: Probability for Applications
Pitman: Probability
Rawlings, Pantula and Dickey: Applied Regression Analysis
Robert: The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, Second Edition
Robert and Casella: Monte Carlo Statistical Methods, Second Edition
Rose and Smith: Mathematical Statistics with *Mathematica*
Ruppert: Statistics and Finance: An Introduction
Santner and Duffy: The Statistical Analysis of Discrete Data
Saville and Wood: Statistical Methods: The Geometric Approach
Sen and Srivastava: Regression Analysis: Theory, Methods, and Applications
Shao: Mathematical Statistics, Second Edition
Shorack: Probability for Statisticians
Shumway and Stoffer: Time Series Analysis and Its Applications
Simonoff: Analyzing Categorical Data
Terrell: Mathematical Statistics: A Unified Introduction
Timm: Applied Multivariate Analysis
Toutenburg: Statistical Analysis of Designed Experiments, Second Edition
Wasserman: All of Statistics: A Concise Course in Statistical Inference
Whittle: Probability via Expectation, Fourth Edition
Zacks: Introduction to Reliability Analysis: Probability Models and Statistical Methods

ALSO AVAILABLE FROM SPRINGER!



THEORY OF POINT ESTIMATION

Second Edition

E.L. LEHMANN and **GEORGE CASELLA**

This second, much enlarged edition by Lehmann and Casella of Lehmann's classic text on point estimation maintains the outlook and general style of the first edition. All of the topics are updated. An entirely new chapter on Bayesian and hierarchical Bayesian approaches is provided, and there is much new material on simultaneous estimation. Each chapter concludes with a Notes section, which contains suggestions for further study. The book is a companion volume to the second edition of Lehmann's *Testing Statistical Hypotheses*.

1998/589 PP./HARDCOVER/ISBN 0-387-98502-6
SPRINGER TEXTS IN STATISTICS

ELEMENTS OF LARGE-SAMPLE THEORY

E.L. LEHMANN

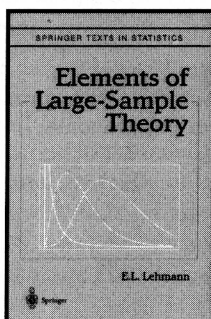
Elements of Large Sample Theory provides a unified treatment of first-order large-sample theory. It discusses a broad range of applications including introductions to density estimation, the bootstrap, and the asymptotics of survey methodology written at an elementary level. The book is suitable for students at the Master's level in statistics and in applied fields who have a background of two years of calculus.

From a review:

"The book also contains rich collection of problems and a useful list of references, and can be warmly recommended as a complementary text to lectures on mathematical statistics, as well as a textbook for more advanced courses."

— EUROPEAN MATHEMATICAL SOCIETY

1999/631 PAGES/HARDCOVER/ISBN 0-387-98595-6
SPRINGER TEXTS IN STATISTICS



SUBSAMPLING

DIMITRIS N. POLITIS, JOSEPH P. ROMANO and
MICHAEL WOLF

Since Efron's profound paper on the bootstrap, an enormous amount of effort has been spent on the development of bootstrap, jackknife, and other resampling methods. The primary goal of these computer-intensive methods has been to provide statistical tools that work in complex situations without imposing unrealistic or unverifiable assumptions about the data generating mechanism. This book sets out to lay some of the foundations for subsampling methodology and related methods.

From a review:

"The book is one of the most comprehensive texts in the subsampling realm and provides a solid background for researchers working in the related areas of statistics."

— V.V. FEDOROV IN "SHORT BOOK REVIEWS", VOL. 21/1, APRIL 2001

1999/347 PP./HARDCOVER/ISBN 0-387-98854-8
SPRINGER SERIES IN STATISTICS

To Order or for Information:

In the Americas: **CALL:** 1-800-SPRINGER or
FAX: (201) 348-4505 • **WRITE:** Springer, Dept.
S7873, P.O. Box 2485, Secaucus, NJ 07096-2485
• **VISIT:** Your local technical bookstore
• **E-MAIL:** orders@springer-ny.com

Outside the Americas: **CALL:** +49 (0) 6221 345-217/8
• **FAX:** + 49 (0) 6221 345-229 • **WRITE:** Springer
Customer Service, Haberstrasse 7, 69126
Heidelberg, Germany • **E-MAIL:** orders@springer.de

PROMOTION: S7873



Springer

the language of science

Contents

| | |
|---|------------|
| Preface | vii |
| I Small-Sample Theory | 1 |
| 1 The General Decision Problem | 3 |
| 1.1 Statistical Inference and Statistical Decisions | 3 |
| 1.2 Specification of a Decision Problem | 4 |
| 1.3 Randomization; Choice of Experiment | 8 |
| 1.4 Optimum Procedures | 9 |
| 1.5 Invariance and Unbiasedness | 11 |
| 1.6 Bayes and Minimax Procedures | 14 |
| 1.7 Maximum Likelihood | 16 |
| 1.8 Complete Classes | 17 |
| 1.9 Sufficient Statistics | 18 |
| 1.10 Problems | 21 |
| 1.11 Notes | 27 |
| 2 The Probability Background | 28 |
| 2.1 Probability and Measure | 28 |
| 2.2 Integration | 31 |
| 2.3 Statistics and Subfields | 34 |
| 2.4 Conditional Expectation and Probability | 36 |
| 2.5 Conditional Probability Distributions | 41 |
| 2.6 Characterization of Sufficiency | 44 |
| 2.7 Exponential Families | 46 |

| | | |
|----------|--|------------|
| 2.8 | Problems | 50 |
| 2.9 | Notes | 55 |
| 3 | Uniformly Most Powerful Tests | 56 |
| 3.1 | Stating The Problem | 56 |
| 3.2 | The Neyman–Pearson Fundamental Lemma | 59 |
| 3.3 | p -values | 63 |
| 3.4 | Distributions with Monotone Likelihood Ratio | 65 |
| 3.5 | Confidence Bounds | 72 |
| 3.6 | A Generalization of the Fundamental Lemma | 77 |
| 3.7 | Two-Sided Hypotheses | 81 |
| 3.8 | Least Favorable Distributions | 83 |
| 3.9 | Applications to Normal Distributions | 86 |
| 3.9.1 | Univariate Normal Models | 86 |
| 3.9.2 | Multivariate Normal Models | 89 |
| 3.10 | Problems | 92 |
| 3.11 | Notes | 107 |
| 4 | Unbiasedness: Theory and First Applications | 110 |
| 4.1 | Unbiasedness For Hypothesis Testing | 110 |
| 4.2 | One-Parameter Exponential Families | 111 |
| 4.3 | Similarity and Completeness | 115 |
| 4.4 | UMP Unbiased Tests for Multiparameter Exponential Families | 119 |
| 4.5 | Comparing Two Poisson or Binomial Populations | 124 |
| 4.6 | Testing for Independence in a 2×2 Table | 127 |
| 4.7 | Alternative Models for 2×2 Tables | 130 |
| 4.8 | Some Three-Factor Contingency Tables | 132 |
| 4.9 | The Sign Test | 135 |
| 4.10 | Problems | 139 |
| 4.11 | Notes | 149 |
| 5 | Unbiasedness: Applications to Normal Distributions | 150 |
| 5.1 | Statistics Independent of a Sufficient Statistic | 150 |
| 5.2 | Testing the Parameters of a Normal Distribution | 153 |
| 5.3 | Comparing the Means and Variances of Two Normal Distribu- tions | 157 |
| 5.4 | Confidence Intervals and Families of Tests | 161 |
| 5.5 | Unbiased Confidence Sets | 164 |
| 5.6 | Regression | 168 |
| 5.7 | Bayesian Confidence Sets | 171 |
| 5.8 | Permutation Tests | 176 |
| 5.9 | Most Powerful Permutation Tests | 177 |
| 5.10 | Randomization As A Basis For Inference | 181 |
| 5.11 | Permutation Tests and Randomization | 184 |
| 5.12 | Randomization Model and Confidence Intervals | 187 |
| 5.13 | Testing for Independence in a Bivariate Normal Distribution | 190 |
| 5.14 | Problems | 192 |
| 5.15 | Notes | 210 |

| | |
|---|------------|
| 6 Invariance | 212 |
| 6.1 Symmetry and Invariance | 212 |
| 6.2 Maximal Invariants | 214 |
| 6.3 Most Powerful Invariant Tests | 218 |
| 6.4 Sample Inspection by Variables | 223 |
| 6.5 Almost Invariance | 225 |
| 6.6 Unbiasedness and Invariance | 229 |
| 6.7 Admissibility | 232 |
| 6.8 Rank Tests | 239 |
| 6.9 The Two-Sample Problem | 242 |
| 6.10 The Hypothesis of Symmetry | 246 |
| 6.11 Equivariant Confidence Sets | 248 |
| 6.12 Average Smallest Equivariant Confidence Sets | 251 |
| 6.13 Confidence Bands for a Distribution Function | 255 |
| 6.14 Problems | 257 |
| 6.15 Notes | 276 |
| 7 Linear Hypotheses | 277 |
| 7.1 A Canonical Form | 277 |
| 7.2 Linear Hypotheses and Least Squares | 281 |
| 7.3 Tests of Homogeneity | 285 |
| 7.4 Two-Way Layout: One Observation per Cell | 287 |
| 7.5 Two-Way Layout: m Observations Per Cell | 290 |
| 7.6 Regression | 293 |
| 7.7 Random-Effects Model: One-way Classification | 297 |
| 7.8 Nested Classifications | 300 |
| 7.9 Multivariate Extensions | 304 |
| 7.10 Problems | 306 |
| 7.11 Notes | 317 |
| 8 The Minimax Principle | 319 |
| 8.1 Tests with Guaranteed Power | 319 |
| 8.2 Examples | 322 |
| 8.3 Comparing Two Approximate Hypotheses | 326 |
| 8.4 Maximin Tests and Invariance | 329 |
| 8.5 The Hunt–Stein Theorem | 331 |
| 8.6 Most Stringent Tests | 337 |
| 8.7 Problems | 338 |
| 8.8 Notes | 347 |
| 9 Multiple Testing and Simultaneous Inference | 348 |
| 9.1 Introduction and the FWER | 348 |
| 9.2 Maximin Procedures | 354 |
| 9.3 The Hypothesis of Homogeneity | 363 |
| 9.4 Scheffé’s S -Method: A Special Case | 375 |
| 9.5 Scheffé’s S -Method for General Linear Models | 380 |
| 9.6 Problems | 385 |
| 9.7 Notes | 391 |

| | |
|--|----------------|
| 10 Conditional Inference | 392 |
| 10.1 Mixtures of Experiments | 392 |
| 10.2 Ancillary Statistics | 395 |
| 10.3 Optimal Conditional Tests | 400 |
| 10.4 Relevant Subsets | 404 |
| 10.5 Problems | 409 |
| 10.6 Notes | 414 |
| II Large-Sample Theory | 417 |
| 11 Basic Large Sample Theory | 419 |
| 11.1 Introduction | 419 |
| 11.2 Basic Convergence Concepts | 424 |
| 11.2.1 Weak Convergence and Central Limit Theorems . . . | 424 |
| 11.2.2 Convergence in Probability and Applications | 431 |
| 11.2.3 Almost Sure Convergence | 440 |
| 11.3 Robustness of Some Classical Tests | 444 |
| 11.3.1 Effect of Distribution | 444 |
| 11.3.2 Effect of Dependence | 448 |
| 11.3.3 Robustness in Linear Models | 451 |
| 11.4 Nonparametric Mean | 459 |
| 11.4.1 Edgeworth Expansions | 459 |
| 11.4.2 The t -test | 462 |
| 11.4.3 A Result of Bahadur and Savage | 466 |
| 11.4.4 Alternative Tests | 468 |
| 11.5 Problems | 469 |
| 11.6 Notes | 480 |
| 12 Quadratic Mean Differentiable Families | 482 |
| 12.1 Introduction | 482 |
| 12.2 Quadratic Mean Differentiability (q.m.d.) | 482 |
| 12.3 Contiguity | 492 |
| 12.4 Likelihood Methods in Parametric Models | 503 |
| 12.4.1 Efficient Likelihood Estimation | 504 |
| 12.4.2 Wald Tests and Confidence Regions | 508 |
| 12.4.3 Rao Score Tests | 511 |
| 12.4.4 Likelihood Ratio Tests | 513 |
| 12.5 Problems | 517 |
| 12.6 Notes | 525 |
| 13 Large Sample Optimality | 527 |
| 13.1 Testing Sequences, Metrics, and Inequalities | 527 |
| 13.2 Asymptotic Relative Efficiency | 534 |
| 13.3 AUMP Tests in Univariate Models | 540 |
| 13.4 Asymptotically Normal Experiments | 549 |
| 13.5 Applications to Parametric Models | 553 |
| 13.5.1 One-sided Hypotheses | 553 |
| 13.5.2 Equivalence Hypotheses | 559 |

| | | |
|-----------|---|------------|
| 13.5.3 | Multi-sided Hypotheses | 564 |
| 13.6 | Applications to Nonparametric Models | 567 |
| 13.6.1 | Nonparametric Mean | 567 |
| 13.6.2 | Nonparametric Testing of Functionals | 570 |
| 13.7 | Problems | 574 |
| 13.8 | Notes | 582 |
| 14 | Testing Goodness of Fit | 583 |
| 14.1 | Introduction | 583 |
| 14.2 | The Kolmogorov-Smirnov Test | 584 |
| 14.2.1 | Simple Null Hypothesis | 584 |
| 14.2.2 | Extensions of the Kolmogorov-Smirnov Test | 589 |
| 14.3 | Pearson's Chi-squared Statistic | 590 |
| 14.3.1 | Simple Null Hypothesis | 590 |
| 14.3.2 | Chi-squared Test of Uniformity | 594 |
| 14.3.3 | Composite Null Hypothesis | 597 |
| 14.4 | Neyman's Smooth Tests | 599 |
| 14.4.1 | Fixed k Asymptotics | 601 |
| 14.4.2 | Neyman's Smooth Tests With Large k | 603 |
| 14.5 | Weighted Quadratic Test Statistics | 607 |
| 14.6 | Global Behavior of Power Functions | 616 |
| 14.7 | Problems | 622 |
| 14.8 | Notes | 629 |
| 15 | General Large Sample Methods | 631 |
| 15.1 | Introduction | 631 |
| 15.2 | Permutation and Randomization Tests | 632 |
| 15.2.1 | The Basic Construction | 632 |
| 15.2.2 | Asymptotic Results | 636 |
| 15.3 | Basic Large Sample Approximations | 643 |
| 15.3.1 | Pivotal Method | 644 |
| 15.3.2 | Asymptotic Pivotal Method | 646 |
| 15.3.3 | Asymptotic Approximation | 647 |
| 15.4 | Bootstrap Sampling Distributions | 648 |
| 15.4.1 | Introduction and Consistency | 648 |
| 15.4.2 | The Nonparametric Mean | 653 |
| 15.4.3 | Further Examples | 655 |
| 15.4.4 | Stepdown Multiple Testing | 658 |
| 15.5 | Higher Order Asymptotic Comparisons | 661 |
| 15.6 | Hypothesis Testing | 668 |
| 15.7 | Subsampling | 673 |
| 15.7.1 | The Basic Theorem in the I.I.D. Case | 674 |
| 15.7.2 | Comparison with the Bootstrap | 677 |
| 15.7.3 | Hypothesis Testing | 680 |
| 15.8 | Problems | 682 |
| 15.9 | Notes | 690 |
| A | Auxiliary Results | 692 |
| A.1 | Equivalence Relations; Groups | 692 |

| | | |
|----------------------|---|------------|
| A.2 | Convergence of Functions; Metric Spaces | 693 |
| A.3 | Banach and Hilbert Spaces | 696 |
| A.4 | Dominated Families of Distributions | 698 |
| A.5 | The Weak Compactness Theorem | 700 |
| References | | 702 |
| Author Index | | 757 |
| Subject Index | | 767 |

Part I

Small-Sample Theory