

# Bootstrap Methods: A Guide for Practitioners and Researchers

Second Edition

MICHAEL R. CHERNICK

United BioSource Corporation  
Newtown, PA



**WILEY-INTERSCIENCE**

A JOHN WILEY & SONS, INC., PUBLICATION



## Bootstrap Methods



---

THE WILEY BICENTENNIAL—KNOWLEDGE FOR GENERATIONS

---

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

**WILLIAM J. PESCE**  
PRESIDENT AND CHIEF EXECUTIVE OFFICER

**PETER BOOTH WILEY**  
CHAIRMAN OF THE BOARD

---

# Bootstrap Methods: A Guide for Practitioners and Researchers

Second Edition

MICHAEL R. CHERNICK

United BioSource Corporation  
Newtown, PA



**WILEY-INTERSCIENCE**

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2008 by John Wiley & Sons, Inc. All rights reserved

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 or the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

Wiley Bicentennial Logo: Richard J. Pacifico

***Library of Congress Cataloging-in-Publication Data:***

Chernick, Michael R.

Bootstrap methods : a guide for practitioners and researchers /

Michael R. Chernick.—2nd ed.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-471-75621-7 (cloth)

1. Bootstrap (Statistics) I. Title.

QA276.8.C48 2008

519.5'44—dc22

2007029309

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

# Contents

<b>Preface to Second Edition</b>	<b>ix</b>
<b>Preface to First Edition</b>	<b>xiii</b>
<b>Acknowledgments</b>	<b>xvii</b>
<b>1. What Is Bootstrapping?</b>	<b>1</b>
1.1. Background, 1	
1.2. Introduction, 8	
1.3. Wide Range of Applications, 13	
1.4. Historical Notes, 16	
1.5. Summary, 24	
<b>2. Estimation</b>	<b>26</b>
2.1. Estimating Bias, 26	
2.1.1. How to Do It by Bootstrapping, 26	
2.1.2. Error Rate Estimation in Discrimination, 28	
2.1.3. Error Rate Estimation: An Illustrative Problem, 39	
2.1.4. Efron's Patch Data Example, 44	
2.2. Estimating Location and Dispersion, 46	
2.2.1. Means and Medians, 47	
2.2.2. Standard Errors and Quartiles, 48	
2.3. Historical Notes, 51	
<b>3. Confidence Sets and Hypothesis Testing</b>	<b>53</b>
3.1. Confidence Sets, 55	
3.1.1. Typical Value Theorems for M-Estimates, 55	
3.1.2. Percentile Method, 57	

3.1.3.	Bias Correction and the Acceleration Constant, 58	
3.1.4.	Iterated Bootstrap, 61	
3.1.5.	Bootstrap Percentile $t$ Confidence Intervals, 64	
3.2.	Relationship Between Confidence Intervals and Tests of Hypotheses, 64	
3.3.	Hypothesis Testing Problems, 66	
3.3.1.	Tendrill DX Lead Clinical Trial Analysis, 67	
3.4.	An Application of Bootstrap Confidence Intervals to Binary Dose–Response Modeling, 71	
3.5.	Historical Notes, 75	
<b>4.</b>	<b>Regression Analysis</b>	<b>78</b>
4.1.	Linear Models, 82	
4.1.1.	Gauss–Markov Theory, 83	
4.1.2.	Why Not Just Use Least Squares? 83	
4.1.3.	Should I Bootstrap the Residuals from the Fit? 84	
4.2.	Nonlinear Models, 86	
4.2.1.	Examples of Nonlinear Models, 87	
4.2.2.	A Quasi-optical Experiment, 89	
4.3.	Nonparametric Models, 93	
4.4.	Historical Notes, 94	
<b>5.</b>	<b>Forecasting and Time Series Analysis</b>	<b>97</b>
5.1.	Methods of Forecasting, 97	
5.2.	Time Series Models, 98	
5.3.	When Does Bootstrapping Help with Prediction Intervals? 99	
5.4.	Model-Based Versus Block Resampling, 103	
5.5.	Explosive Autoregressive Processes, 107	
5.6.	Bootstrapping-Stationary Arma Models, 108	
5.7.	Frequency-Based Approaches, 108	
5.8.	Sieve Bootstrap, 110	
5.9.	Historical Notes, 111	
<b>6.</b>	<b>Which Resampling Method Should You Use?</b>	<b>114</b>
6.1.	Related Methods, 115	
6.1.1.	Jackknife, 115	
6.1.2.	Delta Method, Infinitesimal Jackknife, and Influence Functions, 116	
6.1.3.	Cross-Validation, 119	
6.1.4.	Subsampling, 119	

6.2.	Bootstrap Variants, 120	
6.2.1.	Bayesian Bootstrap, 121	
6.2.2.	The Smoothed Bootstrap, 123	
6.2.3.	The Parametric Bootstrap, 124	
6.2.4.	Double Bootstrap, 125	
6.2.5.	The m-out-of-n Bootstrap, 125	
<b>7.</b>	<b>Efficient and Effective Simulation</b>	<b>127</b>
7.1.	How Many Replications? 128	
7.2.	Variance Reduction Methods, 129	
7.2.1.	Linear Approximation, 129	
7.2.2.	Balanced Resampling, 131	
7.2.3.	Antithetic Variates, 132	
7.2.4.	Importance Sampling, 133	
7.2.5.	Centering, 134	
7.3.	When Can Monte Carlo Be Avoided? 135	
7.4.	Historical Notes, 136	
<b>8.</b>	<b>Special Topics</b>	<b>9</b>
8.1.	Spatial Data, 139	
8.1.1.	Kriging, 139	
8.1.2.	Block Bootstrap on Regular Grids, 142	
8.1.3.	Block Bootstrap on Irregular Grids, 143	
8.2.	Subset Selection, 143	
8.3.	Determining the Number of Distributions in a Mixture Model, 145	
8.4.	Censored Data, 148	
8.5.	<i>p</i> -Value Adjustment, 149	
8.5.1.	Description of Westfall–Young Approach, 150	
8.5.2.	Passive Plus DX Example, 150	
8.5.3.	Consulting Example, 152	
8.6.	Bioequivalence Applications, 153	
8.6.1.	Individual Bioequivalence, 153	
8.6.2.	Population Bioequivalence, 155	
8.7.	Process Capability Indices, 156	
8.8.	Missing Data, 164	
8.9.	Point Processes, 166	
8.10.	Lattice Variables, 168	
8.11.	Historical Notes, 169	

<b>9. When Bootstrapping Fails Along with Remedies for Failures</b>	<b>172</b>
9.1. Too Small of a Sample Size, 173	
9.2. Distributions with Infinite Moments, 175	
9.2.1. Introduction, 175	
9.2.2. Example of Inconsistency, 176	
9.2.3. Remedies, 176	
9.3. Estimating Extreme Values, 177	
9.3.1. Introduction, 177	
9.3.2. Example of Inconsistency, 177	
9.3.3. Remedies, 178	
9.4. Survey Sampling, 179	
9.4.1. Introduction, 179	
9.4.2. Example of Inconsistency, 180	
9.4.3. Remedies, 180	
9.5. Data Sequences that Are $M$ -Dependent, 180	
9.5.1. Introduction, 180	
9.5.2. Example of Inconsistency When Independence Is Assumed, 181	
9.5.3. Remedies, 181	
9.6. Unstable Autoregressive Processes, 182	
9.6.1. Introduction, 182	
9.6.2. Example of Inconsistency, 182	
9.6.3. Remedies, 183	
9.7. Long-Range Dependence, 183	
9.7.1. Introduction, 183	
9.7.2. Example of Inconsistency, 183	
9.7.3. Remedies, 184	
9.8. Bootstrap Diagnostics, 184	
9.9. Historical Notes, 185	
<b>Bibliography 1 (Prior to 1999)</b>	<b>188</b>
<b>Bibliography 2 (1999–2007)</b>	<b>274</b>
<b>Author Index</b>	<b>330</b>
<b>Subject Index</b>	<b>359</b>

# Preface to Second Edition

Since the publication of the first edition of this book in 1999, there have been many additional and important applications in the biological sciences as well as in other fields. The major theoretical and applied books have not yet been revised. They include Hall (1992a), Efron and Tibshirani (1993), Hjorth (1994), Shao and Tu (1995), and Davison and Hinkley (1997). In addition, the bootstrap is being introduced much more often in both elementary and advanced statistics books—including Chernick and Friis (2002), which is an example of an elementary introductory biostatistics book.

The first edition stood out for (1) its use of some real-world applications not covered in other books and (2) its extensive bibliography and its emphasis on the wide variety of applications. That edition also pointed out instances where the bootstrap principle fails and why it fails. Since that time, additional modifications to the bootstrap have overcome some of the problems such as some of those involving finite populations, heavy-tailed distributions, and extreme values. Additional important references not included in the first edition are added to that bibliography. Many applied papers and other references from the period of 1999–2007 are included in a second bibliography. I did not attempt to make an exhaustive update of references.

The collection of articles entitled *Frontiers in Statistics*, published in 2006 by Imperial College Press as a tribute to Peter Bickel and edited by Jianqing Fan and Hira Koul, contains a section on bootstrapping and statistical learning including two chapters directly related to the bootstrap (Chapter 10, Boosting Algorithms: With an Application to Bootstrapping Multivariate Time Series; and Chapter 11, Bootstrap Methods: A Review). There is some reference to Chapter 10 from *Frontiers in Statistics* which is covered in the expanded Chapter 8, Special Topics; and material from Chapter 11 of *Frontiers in Statistics* will be used throughout the text.

Lahiri, the author of Chapter 11 in *Frontiers in Statistics*, has also published an excellent text on resampling methods for dependent data, Lahiri (2003a), which deals primarily with bootstrapping in dependent situations, particularly time series and spatial processes. Some of this material will be covered in

Chapters 4, 5, 8, and 9 of this text. For time series and other dependent data, the moving block bootstrap has become the method of choice and other block bootstrap methods have been developed. Other bootstrap techniques for dependent data include transformation-based bootstrap (primarily the frequency domain bootstrap) and the sieve bootstrap. Lahiri has been one of the pioneers at developing bootstrap methods for dependent data, and his text Lahiri (2003a) covers these methods and their statistical properties in great detail along with some results for the IID case. To my knowledge, it is the only major bootstrap text with extensive theory and applications from 2001 to 2003.

Since the first edition of my text, I have given a number of short courses on the bootstrap using materials from this and other texts as have others. In the process, new examples and illustrations have been found that are useful in a course text. The bootstrap is also being taught in many graduate school statistics classes as well as in some elementary undergraduate classes. The value of bootstrap methods is now well established.

The intention of the first edition was to provide a historical perspective to the development of the bootstrap, to provide practitioners with enough applications and references to know when and how the bootstrap can be used and to also understand its pitfalls. It had a second purpose to introduce others to the bootstrap, who may not be familiar with it, so that they can learn the basics and pursue further advances, if they are so interested. It was not intended to be used exclusively as a graduate text on the bootstrap. However, it could be used as such with supplemental materials, whereas the text by Davison and Hinkley (1997) is a self-contained graduate-level text. In a graduate course, this book could also be used as supplemental material to one of the other fine texts on bootstrap, particularly Davison and Hinkley (1997) and Efron and Tibshirani (1993). Student exercises were not included; and although the number of illustrative examples is increased in this edition, I do not include exercises at the end of the chapters.

For the most part the first edition was successful, but there were a few critics. The main complaints were with regard to lack of detail in the middle and latter chapters. There, I was sketchy in the exposition and relied on other reference articles and texts for the details. In some cases the material had too much of an encyclopedic flavor. Consequently, I have expanded on the description of the bootstrap approach to censored data in Section 8.4, and to  $p$ -value adjustment in Section 8.5. In addition to the discussion of kriging in Section 8.1, I have added some coverage of other results for spatial data that is also covered in Lahiri (2003a).

There are no new chapters in this edition and I tried not to add too many pages to the original bibliography, while adding substantially to Chapters 4 (on regression), 5 (on forecasting and time series), 8 (special topics), and 9 (when the bootstrap fails and remedies) and somewhat to Chapter 3 (on hypothesis testing and confidence intervals). Applications in the pharmaceutical industry such as the use of bootstrap for estimating individual and population bioequivalence are also included in a new Section 8.6.

Chapter 2 on estimating bias covered the error rate estimation problem in discriminant analysis in great detail. I find no need to expand on that material because in addition to McLachlan (1992), many new books and new editions of older books have been published on statistical pattern recognition, discriminant analysis, and machine learning that include good coverage of the bootstrap application to error rate estimation.

The first edition got mixed reviews in the technical journals. Reviews by bootstrap researchers were generally very favorable, because they recognized the value of consolidating information from diverse sources into one book. They also appreciated the objectives I set for the text and generally felt that the book met them. In a few other reviews from statisticians not very familiar with all the bootstrap applications, who were looking to learn details about the techniques, they wrote that there were too many pages devoted to the bibliography and not enough to exposition of the techniques.

My choice here is to add a second bibliography with references from 1999–2006 and early 2007. This adds about 1000 new references that I found primarily through a simple search of all articles and books with “bootstrap” as a key word or as part of the title, in the *Current Index to Statistics* (CIS) through my online access. For others who have access to such online searches, it is now much easier to find even obscure references as compared to what could be done in 1999 when the first edition of this book came out.

In the spirit of the first edition and in order to help readers who may not have easy access to such internet sources, I have decided to include all these new references in the second bibliography with those articles and books that are cited in the text given asterisks. This second bibliography has the citations listed in order by year of publication (starting with 1999) and in alphabetical order by first author’s last name for each year. This simple addition to the bibliographies nearly doubles the size of the bibliographic section. I have also added more than a dozen references to the old bibliography [now called Bibliography 1 (prior to 1999)] from references during the period from 1985 to 1998 that were not included in the first edition.

To satisfy my critics, I have also added exposition to the chapters that needed it. I hope that I have remedied some of the criticism without sacrificing the unique aspects that some reviewers and many readers found valuable in the first edition.

I believe that in my determination to address the needs of two groups with different interests, I had to make compromises, avoiding a detailed development of theory for the first group and providing a long list of references for the second group that wanted to see the details. To better reflect and emphasize the two groups that the text is aimed at, I have changed the subtitle from *A Practitioner’s Guide* to *A Guide for Practitioners and Researchers*. Also, because of the many remedies that have been devised to overcome the failures of the bootstrap and because I also include some remedies along with the failures, I have changed the title of Chapter 9 from “When does Bootstrapping

Fail?” to “When Bootstrapping Fails Along with Some Remedies for Failures.”

The bibliography also was intended to help bootstrap specialists become aware of other theoretical and applied work that might appear in journals that they do not read. For them this feature may help them to be abreast of the latest advances and thus be better prepared and motivated to add to the research.

This compromise led some from the first group to feel overwhelmed by technical discussion, wishing to see more applications and not so many pages of references that they probably will never look at. For the second group, the bibliography is better appreciated but there is a desire to see more pages devoted to exposition of the theory and greater detail to the theory and more pages for applications (perhaps again preferring more pages in the text and less in the bibliography). While I did continue to expand the bibliographic section of the book, I do hope that the second edition will appeal to the critics in both groups by providing additional applications and more detailed and clear exposition of the methodology. I also hope that they will not mind the two extensive bibliographies that make my book the largest single source for extensive references on bootstrap.

Although somewhat out of date, the preface to the first edition still provides a good description of the goals of the book and how the text compares to some of its main competitors. Only objective 5 in that preface was modified. With the current state of the development of websites on the internet, it is now very easy for almost anyone to find these references online through the use of sophisticated search engines such as Yahoo’s or Google’s or through a CIS search.

I again invite readers to notify me of any errors or omissions in the book. There continue to be many more papers listed in the bibliographies than are referenced in the text. In order to make clear which references are cited in the text, I put an asterisk next to the cited references but I now have dispensed with a numbering according to alphabetical order, which only served to give a count of the number of books and articles cited in the text.

*United BioSource Corporation  
Newtown, Pennsylvania  
July 2007*

MICHAEL R. CHERNICK

# Preface to First Edition

The bootstrap is a resampling procedure. It is named that because it involves resampling from the original data set. Some resampling procedures similar to the bootstrap go back a long way. The use of computers to do simulation goes back to the early days of computing in the late 1940s. However, it was Efron (1979a) that unified ideas and connected the simple nonparametric bootstrap, which “resamples the data with replacement,” with earlier accepted statistical tools for estimating standard errors, such as the jackknife and the delta method.

The purpose of this book is to (1) provide an introduction to the bootstrap for readers who do not have an advanced mathematical background, (2) update some of the material in the Efron and Tibshirani (1993) book by presenting results on improved confidence set estimation, estimation of error rates in discriminant analysis, and applications to a wide variety of hypothesis testing and estimation problems, (3) exhibit counterexamples to the consistency of bootstrap estimates so that the reader will be aware of the limitations of the methods, (4) connect it with some older and more traditional resampling methods including the permutation tests described by Good (1994), and (5) provide a bibliography that is extensive on the bootstrap and related methods up through 1992 with key additional references from 1993 through 1998, including new applications.

The objectives of the book are very similar to those of Davison and Hinkley (1997), especially (1) and (2). However, I differ in that this book does not contain exercises for students, but it does include a much more extensive bibliography.

This book is not a classroom text. It is intended to be a reference source for statisticians and other practitioners of statistical methods. It could be used as a supplement on an undergraduate or graduate course on resampling methods for an instructor who wants to incorporate some real-world applications and supply additional motivation for the students.

The book is aimed at an audience similar to the one addressed by Efron and Tibshirani (1993) and does not develop the theory and mathematics to

the extent of Davison and Hinkley (1997). Mooney and Duval (1993) and Good (1998) are elementary accounts, but they do not provide enough development to help the practitioner gain a great deal of insight into the methods.

The spectacular success of the bootstrap in error rate estimation for discriminant functions with small training sets along with my detailed knowledge of the subject justifies the extensive coverage given to this topic in Chapter 2. A text that provides a detailed treatment of the classification problem and is the only text to include a comparison of bootstrap error rate estimates with other traditional methods is McLachlan (1992).

Mine is the first text to provide extensive coverage of real-world applications for practitioners in many diverse fields. I also provide the most detailed guide yet available to the bootstrap literature. This I hope will motivate research statisticians to make theoretical and applied advances in bootstrapping.

Several books (at least 30) deal in part with the bootstrap in specific contexts, but none of these are totally dedicated to the subject [Sprent (1998) devotes Chapter 2 to the bootstrap and provides discussion of bootstrap methods throughout his book]. Schervish (1995) provides an introductory discussion on the bootstrap in Section 5.3 and cites Young (1994) as an article that provides a good overview of the subject. Babu and Feigelson (1996) address applications of statistics in astronomy. They refer to the statistics of astronomy as astrostatistics. Chapter 5 (pp. 93–103) of the Babu–Feigelson text covers resampling methods emphasizing the bootstrap. At this point there are about a half dozen other books devoted to the bootstrap, but of these only four (Davison and Hinkley, 1997; Manly, 1997; Hjorth, 1994; Efron and Tibshirani, 1993) are not highly theoretical.

Davison and Hinkley (1997) give a good account of the wide variety of applications and provide a coherent account of the theoretical literature. They do not go into the mathematical details to the extent of Shao and Tu (1995) or Hall (1992a). Hjorth (1994) is unique in that it provides detailed coverage of model selection applications.

Although many authors are now including the bootstrap as one of the tools in a statistician's arsenal (or for that matter in the tool kit of any practitioner of statistical methods), they deal with very specific applications and do not provide a guide to the variety of uses and the limitations of the techniques for the practitioner. This book is intended to present the practitioner with a guide to the use of the bootstrap while at the same time providing him or her with an awareness of its known current limitations. As an additional bonus, I provide an extensive guide to the research literature on the bootstrap.

This book is aimed at two audiences. The first consists of applied statisticians, engineers, scientists, and clinical researchers who need to use statistics in their work. For them, I have tried to maintain a low mathematical level. Consequently, I do not go into the details of stochastic convergence or the Edgeworth and Cornish–Fisher expansions that are important in determining

the rate of convergence for various estimators and thus identify the higher-order efficiency of some of these estimators and the properties of their approximate confidence intervals.

However, I do not avoid discussion of these topics. Readers should bear with me. There is a need to understand the role of these techniques and the corresponding bootstrap theory in order to get an appreciation and understanding of how, why, and when the bootstrap works. This audience should have some background in statistical methods (at least having completed one elementary statistics course), but they need not have had courses in calculus, advanced mathematics, advanced probability, or mathematical statistics.

The second primary audience is the mathematical statistician who has done research in statistics but has not become familiar with the bootstrap but wants to learn more about it and possibly use it in future research. For him or her, my historical notes and extensive references to applications and theoretical papers will be helpful. This second audience may also appreciate the way I try to tie things together with a somewhat objective view.

To a lesser extent a third group, the serious bootstrap researcher, may find value in this book and the bibliography in particular. I do attempt to maintain technical accuracy, and the bibliography is extensive with many applied papers that may motivate further research. It is more extensive than one obtained simply by using the key word search for “bootstrap” and “resampling” in the *Current Index to Statistics* CD ROM. However, I would not try to claim that such a search could not uncover at least a few articles that I may have missed.

I invite readers to notify me of any errors or omissions in the book, particularly omissions regarding references. There are many more papers listed in the bibliography than are referenced in the text. In order to make clear which references are cited in the text, I put an asterisk next to the cited references along with a numbering according to alphabetical order.

*Diamond Bar, California*  
*January 1999*

MICHAEL R. CHERNICK



# Acknowledgments

When the first edition was written, Peter Hall was kind enough to send an advance copy of his book *The Bootstrap and Edgeworth Expansion* (Hall, 1992a), which was helpful to me especially in explaining the virtues of the various forms of bootstrap confidence intervals. Peter has been a major contributor to various branches of probability and statistics and has been and continues to be a major contributor to bootstrap theory and methods. I have learned a great deal about bootstrapping from Peter and his student Michael Martin, from Peter's book, and from his many papers with Martin and others.

Brad Efron taught me mathematical statistics when I was a graduate student at Stanford. I learned about some of the early developments in bootstrapping first hand from him as he was developing his early ideas on the bootstrap. To me he was a great teacher, mentor, and later a colleague. Although I did not do my dissertation work with him and did not do research on the bootstrap until several years after my graduation, he always encouraged me and gave me excellent advice through many discussions at conferences and seminars and through our various private communications. My letters to him tended to be long and complicated. His replies to me were always brief but right to the point and very helpful. His major contributions to statistical theory include the geometry of exponential families, empirical Bayes methods, and of course the bootstrap. He also has applied the theory to numerous applications in diverse fields. Even today he is publishing important work on microarray data and applications of statistics in physics and other hard sciences. He originated the nonparametric bootstrap and developed many of its properties through the use of Monte Carlo approximations to bootstrap estimates in simulation studies. The Monte Carlo approximation provides a very practical way to use the computer to attain these estimates. Efron's work is evident throughout this text.

This book was originally planned to be half of a two-volume series on resampling methods that Phillip Good and I started. Eventually we decided to publish separate books. Phil has since published three editions to his book,

and this is the second edition of mine. Phil was very helpful to me in organizing the chapter subjects and proofreading many of my early chapters. He continually reminded me to bring out the key points first.

This book started as a bibliography that I was putting together on bootstrap in the early 1990s. The bibliography grew as I discovered, through a discussion with Brad Efron, that Joe Romano and Michael Martin also had been doing a similar thing. They graciously sent me what they had and I combined it with mine to create a large and growing bibliography that I had to continually update throughout the 1990s to keep it current and as complete as possible. Just prior to the publication of the first edition, I used the services of NERAC, a literature search firm. They found several articles that I had missed, particularly those articles that appeared in various applied journals during the period from 1993 through 1998. Gerri Beth Potash of NERAC was the key person who helped with the search. Also, Professor Robert Newcomb from the University of California at Irvine helped me search through an electronic version of the *Current Index to Statistics*. He and his staff at the UCI Statistical Consulting Center (especially Mira Hornbacher) were very helpful with a few other search requests that added to what I obtained from NERAC.

I am indebted to the many typists who helped produce numerous versions of the first edition. The list includes Sally Murray from Nichols Research Corporation, Cheryl Larsson from UC Irvine, and Jennifer Del Villar from Pacesetter. For the second edition I got some help learning about Latex and received guidance and encouragement from my editor Steve Quigley, Susanne Steitz and Jackie Palmieri of the Wiley editorial staff. Sue Hobson from Auxilium was also helpful to me in my preparation of the revised manuscript. However, the typing of the manuscript for the second edition is mine and I am responsible for any typos.

My wife Ann has been a real trooper. She helped me through my illness and allowed me the time to complete the first edition during a very busy period because my two young sons were still preschoolers. She encouraged me to finish the first edition and has been accommodating to my needs as I prepared the second. I do get the common question “Why haven’t you taken out the garbage yet?” My pat answer to that is “Later, I have to finish some work on the book first!” I must thank her for patience and perseverance.

The boys, Daniel and Nicholas, are now teenagers and are much more self-sufficient. My son Nicholas is so adept with computers now that he was able to download improved software for the word processing on my home computer.

## What Is Bootstrapping?

### 1.1. BACKGROUND

The bootstrap is a form of a larger class of methods that resample from the original data set and thus are called resampling procedures. Some resampling procedures similar to the bootstrap go back a long way [e.g., the jackknife goes back to Quenouille (1949), and permutation methods go back to Fisher and Pitman in the 1930s]. Use of computers to do simulation also goes back to the early days of computing in the late 1940s.

However, it was Efron (1979a) who unified ideas and connected the simple nonparametric bootstrap, for independent and identically distributed (IID) observations, which “resamples the data with replacement,” with earlier accepted statistical tools for estimating standard errors such as the jackknife and the delta method. This first method is now commonly called the nonparametric IID bootstrap. It was only after the later papers by Efron and Gong (1983), Efron and Tibshirani (1986), and Diaconis and Efron (1983) and the monograph Efron (1982a) that the statistical and scientific community began to take notice of many of these ideas, appreciate the extensions of the methods and their wide applicability, and recognize their importance.

After the publication of the Efron (1982a) monograph, research activity on the bootstrap grew exponentially. Early on, there were many theoretical developments on the asymptotic consistency of bootstrap estimates. In some of these works, cases where the bootstrap estimate failed to be a consistent estimator for the parameter were uncovered.

Real-world applications began to appear. In the early 1990s the emphasis shifted to finding applications and variants that would work well in practice. In the 1980s along with the theoretical developments, there were many simulation studies that compared the bootstrap and its variants with other competing estimators for a variety of different problems. It also became clear that

although the bootstrap had significant practical value, it also had some limitations.

A special conference of the Institute of Mathematical Statistics was held in Ann Arbor Michigan in May 1990, where many of the prominent bootstrap researchers presented papers exploring the applications and limitations of the bootstrap. The proceedings of this conference were compiled in the book *Exploring the Limits of Bootstrap*, edited by LePage and Billard and published by Wiley in 1992.

A second similar conference, also held in 1990 in Tier, Germany, covered many developments in bootstrapping. The European conference covered Monte Carlo methods, bootstrap confidence bands and prediction intervals, hypothesis tests, time series methods, linear models, special topics, and applications. Limitations of the methods were not addressed at this conference. Its proceedings were published in 1992 by Springer-Verlag. The editors for the proceedings were Jöckel, Rothe, and Sendler.

Although Efron introduced his version of the bootstrap in a 1977 Stanford University Technical Report [later published in a well-known paper in the *Annals of Statistics* (Efron, 1979a)], the procedure was slow to catch on. Many of the applications only began to be covered in textbooks in the 1990s.

Initially, there was a great deal of skepticism and distrust regarding bootstrap methodology. As mentioned in Davison and Hinkley (1997, p. 3): “In the simplest nonparametric problems, we do literally sample from the data, and a common initial reaction is that this is a fraud. In fact it is not.” The article in *Scientific American* (Diaconis and Efron, 1983) was an attempt to popularize the bootstrap in the scientific community by explaining it in layman’s terms and exhibiting a variety of important applications. Unfortunately, by making the explanation simple, technical details were glossed over and the article tended to increase the skepticism rather than abate it.

Other efforts to popularize the bootstrap that were partially successful with the statistical community were Efron (1982a), Efron and Gong (1981), Efron and Gong (1983), Efron (1979b), and Efron and Tibshirani (1986). Unfortunately it was only the *Scientific American* article that got significant exposure to a wide audience of scientists and researchers.

While working at the Aerospace Corporation in the period from 1980 to 1988, I observed that because of the *Scientific American* article, many of the scientist and engineers that I worked with had misconceptions about the methodology. Some supported it because they saw it as a way to use simulation in place of additional sampling (a misunderstanding of what kind of information the Monte Carlo approximation to the bootstrap actually gives). Others rejected it because they interpreted the *Scientific American* article as saying that the technique allowed inferences to be made from data without assumptions by replacing the need for additional “real” data with “simulated” data, and they viewed this as phony science (this is a misunderstanding that comes about because of the oversimplified exposition in the article).

Both views were expressed by my engineering colleagues at the Aerospace Corporation, and I found myself having to try to dispel both of these notions. In so doing, I got to thinking about how the bootstrap could help me in my own research and I saw there was a need for a book like this one. I also felt that in order for articles or books to popularize bootstrap techniques among the scientist, engineers, and other potential practitioners, some of the mathematical and statistical justification had to be presented and any text that skimmed over this would be doomed for failure.

The monograph by Mooney and Duvall (1993) presents only a little of the theory and in my view fails to provide the researcher with even an intuitive feel for why the methodology works. The text by Efron and Tibshirani (1993) was the first attempt at presenting the general methodology and applications to a broad audience of social scientists and researchers. Although it seemed to me to do a very good job of reaching that broad audience, Efron mentioned that he felt that parts of the text were still a little too technical to be clear to everyone in his intended audience.

There is a fine line to draw between being too technical to be understood by those without a strong mathematical background and being too simple to provide a true picture of the methodology devoid of misconceptions. To explain the methodology to those who do not have the mathematical background for a deep understanding of the bootstrap theory, we must avoid technical details on stochastic convergence and other advanced probability tools. But we cannot simplify it to the extent of ignoring the theory because that leads to misconceptions such as the two main ones previously mentioned.

In the late 1970s when I was a graduate student at Stanford University, I saw the theory develop first-hand. Although I understood the technique, I failed to appreciate its value. I was not alone, since many of my fellow graduate students also failed to recognize its great potential. Some statistics professors were skeptical about its usefulness as an addition to the current parametric, semiparametric, and nonparametric techniques.

Why didn't we give the bootstrap more consideration? At that time the bootstrap seemed so simple and straightforward. We did not see it as a part of a revolution in statistical thinking and approaches to data analysis. But today it is clear that this is exactly what it was!

A second reason why some graduate students at Stanford, and possibly other universities, did not elect the bootstrap as a topic for their dissertation research (including Naihua Duan, who was one of Efron's students at that time) is that the key asymptotic properties of the bootstrap appeared to be very difficult to prove. The mathematical approaches and results only began to be known when the papers by Bickel and Freedman (1981) and Singh (1981) appeared, and this was two to three years after many of us had graduated.

Gail Gong was one of Efron's students and the first Stanford graduate student to do a dissertation on the bootstrap. From that point on, many

students at Stanford and other universities followed as the flood gates opened to bootstrap research. Rob Tibshirani was another graduate student of Efron who did his dissertation research on the bootstrap and followed it up with the statistical science article (Efron and Tibshirani, 1986), a book with Trevor Hastie on general additive models, and the text with Efron on the bootstrap (Efron and Tibshirani, 1993). Other Stanford dissertations on bootstrap were Therneau (1983) and Hesterberg (1988). Both dealt with variance reduction techniques for reducing the number of bootstrap iterations necessary to get the Monte Carlo approximation to the bootstrap estimate to achieve a desired level of accuracy with respect to the bootstrap estimate (which is the limit as the number of bootstrap iterations approaches infinity).

My interest in bootstrap research began in earnest in 1983 after I read Efron's paper (Efron, 1983) on the bias adjustment in error rate estimation for classification problems. This applied directly to some of the work I was doing on target discrimination at the Aerospace Corporation and also later at Nichols Research Corporation. This led to a series of simulation studies that I published with Carlton Nealy and Krishna Murthy.

In the late 1980s I met Phil Good, who is an expert on permutation methods and was looking for a way to solve a particular problem that he was having trouble setting up in the framework of a permutation test. I suggested a straightforward bootstrap approach, and this led to comparisons of various procedures to solve the problem. It also opened up a dialogue between us about the virtues of permutation methods, bootstrap methods and other resampling methods, and the basic conditions for their applicability. We recognized that bootstrap and permutation tests were both part of the various resampling procedures that were becoming so useful but were not taught in the introductory statistics courses. That led him to write a series of books on permutation tests and resampling methods and led me to write the first edition of this text and later to incorporate the bootstrap in an introductory course in biostatistics and the text that Professor Robert Friis and I subsequently put together for the course (Chernick and Friis, 2002).

In addition to both being resampling methods, bootstrap and permutation methods could be characterized as computer-intensive, depending on the application. Both approaches avoid unverified parametric assumptions, by relying solely on the original sample. Both require minimal assumptions such as exchangeability of the observations under the null hypothesis. Exchangeability is a property of a random sample that is slightly weaker than the assumption that observations are independent and identically distributed. To be mathematically formal, for a sequence of  $n$  observations the sequence is exchangeable if the probability distribution of any  $k$  consecutive observations ( $k = 1, 2, 3, \dots, n$ ) does not change when the order of the observations is changed through a permutation.

The importance of the bootstrap is now generally recognized as has been noted in the article in the supplemental volume of the *Encyclopedia of Statistical Sciences* (1989 Bootstrapping—II by David Banks, pp. 17–22), the

inclusion of Efron's 1979 *Annals of Statistics* paper in *Breakthroughs in Statistics*, Volume II: *Methodology and Distribution*, S. Kotz and N. L. Johnson, editors (1992, pp. 565–595 with an introduction by R. Beran), and Hall's 1988 *Annals of Statistics* paper in *Breakthroughs in Statistics*, Volume III, S. Kotz and N. L. Johnson, editors (1997, pp. 489–518 with an introduction by E. Mammen). We can also find the bootstrap referenced prominently in the *Encyclopedia of Biostatistics*, with two entries in Volume I: (1) "Bootstrap Methods" by DeAngelis and Young (1998) and (2) "Bootstrapping in Survival Analysis" by Sauerbrei (1998).

The bibliography in the first edition contained 1650 references, and I have only expanded it as necessary. In the first edition I put an asterisk next to each of the 619 references that were referenced directly in the text and also numbered them in the alphabetical order that they were listed. In this edition I continue to use the asterisk to identify those books and articles referenced directly in the text but no longer number them.

The idea of sampling with replacement from the original data did not begin with Efron. Also even earlier than the first use of bootstrap sampling, there were a few related techniques that are now often referred to as resampling techniques. These other techniques predate Efron's bootstrap. Among them are the jackknife, cross-validation, random subsampling, and permutation procedures. Permutation tests have been addressed in standard books on nonparametric inference and in specialized books devoted exclusively to permutation tests including Good (1994, 2000), Edgington (1980, 1987, 1995), and Manly (1991, 1997).

The idea of resampling from the empirical distribution to form a Monte Carlo approximation to the bootstrap estimate may have been thought of and used prior to Efron. Simon (1969) has been referenced by some to indicate his use of the idea as a tool in teaching elementary statistics prior to Efron. Bruce and Simon have been instrumental in popularizing the bootstrap approach through their company Resampling Stats Inc. and their associated software. They also continue to use the Monte Carlo approximation to the bootstrap as a tool for introducing statistical concepts in a first elementary course in statistics [see Simon and Bruce (1991, 1995)]. Julian Simon died several years ago; but Peter Bruce continues to run the company and in addition to teaching resampling in online courses, he has set up a faculty to teach a variety of online statistics courses.

It is clear, however, that widespread use of the methods (particularly by professional statisticians) along with the many theoretical developments occurred only after Efron's 1979 work. That paper (Efron, 1979a) connected the simple bootstrap idea to established methods for estimating the standard error of an estimator, namely, the jackknife, cross-validation, and the delta method, thus providing the theoretical underpinnings that that were then further developed by Efron and other researchers.

There have been other procedures that have been called bootstrap that differ from Efron's concept. I mention two of them in Section 1.4. Whenever

I refer to the bootstrap in this text, I will be referring to Efron's version. Even Efron's bootstrap has many modifications. Among these are the double bootstrap, the smoothed bootstrap, the parametric bootstrap (discussed in Chapter 6), and the Bayesian bootstrap (which was introduced by Rubin in the missing data application described in Section 8.7). Some of the variants of the bootstrap are discussed in Section 2.1.2, including specialized methods specific to the classification problem [e.g., the 632 estimator introduced in Efron (1983) and the convex bootstrap introduced in Chernick, Murthy, and Nealy (1985)].

In May 1998 a conference was held at Rutgers University, organized by Kesar Singh, a Rutgers statistics professor who is a prominent bootstrap researcher. The purpose of the conference was to provide a collection of papers on recent bootstrap developments by key bootstrap researchers and to celebrate the approximately 20 years of research since Efron's original work [first published as a Stanford Technical Report in 1977 and subsequently in the *Annals of Statistics* (Efron, 1979a)]. Abstracts of the papers presented were available from the Rutgers University Statistics Department web site.

Although no proceedings were published for the conference, I received copies of many of the papers by direct request to the authors. The presenters at the meeting included Michael Sherman, Brad Efron, Gutti Babu, C. R. Rao, Kesar Singh, Alastair Young, Dimitris Politis, J.-J. Ren, and Peter Hall. The papers that I received are included in the bibliography. They are Babu, Pathak, and Rao (1998), Sherman and Carlstein (1997), Efron and Tibshirani (1998), and Babu (1998).

This book is organized as follows. Chapter 1 introduces the key ideas and describes the wide range of applications. Chapter 2 deals with estimation and particularly the bias-adjusted estimators with emphasis on error rate estimation for discriminant functions. It shows through simulation studies how the bootstrap and variants such as the 632 estimator perform compared to the more traditional methods when the number of training samples is small. Also discussed are ratio estimates, estimates of medians, standard errors, and quantiles.

Chapter 3 covers confidence intervals and hypothesis tests. The 1–1 correspondence between confidence intervals and hypothesis tests is used to construct hypothesis tests based on bootstrap confidence intervals. We cover two so-called percentile methods and show how more accurate and correct bootstrap confidence intervals can be constructed. In particular, the hierarchy of percentile methods improved by bias correction BC and then BCa is given along with the rate of convergence for these methods and the weakening assumptions required for the validity of the method.

An application in a clinical trial to demonstrate the efficacy of the Tendril DX steroid lead in comparison to nonsteroid leads is also presented. Also covered is a very recent application to adaptive design clinical trials. In this application, proof of concept along with dose–response model identification methods and minimum effective dose estimates are included based on an

adaptive design. The author uses the MED as a parameter to generate “semi-parametric” bootstrap percentile methods.

Chapter 4 covers regression problems, both linear and nonlinear. An application of bootstrap estimates in nonlinear regression of the standard errors of parameters is given for a quasi-optical experiment. New in this edition is the coverage of bootstrap methods applied to outlier detection in least-squares regression.

Chapter 5 addresses time series models and related forecasting problems. This includes model based bootstrap and the various forms of block bootstrap. At the time of the first edition, the moving block bootstrap had been developed but was not very mature. Over the eight intervening years, there have been additional variations on the block bootstrap and more theory and applications. Recently, these developments have been well summarized in the text Lahiri (2003a). We have included some of those block bootstrap methods as well as the sieve bootstrap.

Chapter 6 provides a comparison with other resampling methods and recommends the preferred approach when there is clear evidence in the literature, either through theory or simulation, of its superiority. This was a unique feature of the book when the first edition was published. We have added to our list of resampling methods the  $m$  out of  $n$  bootstrap that we did not cover in the first edition. Although the  $m$  out of  $n$  bootstrap had been considered as a method to consider, it has only recently been proven to be important as a way to remedy inconsistency problems of the naïve bootstrap in many cases.

Chapter 7 deals with simulation methods, emphasizing the variety of available variance reduction techniques and showing the applications for which they can effectively be applied. This chapter is essentially the same as in the first edition.

Chapter 8 gives an account of a variety of miscellaneous topics. These include kriging (a form of smoothing in the analysis of spatial data) and other applications to spatial data, survey sampling, subset selection in both regression and discriminant analysis, analysis of censored data,  $p$ -value adjustment for multiplicity, estimation of process capability indices (measures of manufacturing process performance in quality assurance work), application of the Bayesian bootstrap in missing data problems, and the estimation of individual and population bioequivalence in pharmaceutical studies (often used to get acceptance of a generic drug when compared to a similar market-approved drug).

Chapter 9 describes examples in the literature where the ordinary bootstrap procedures fail. In many instances, modifications have been devised to overcome the problem, and these are discussed. In the first edition, remedies for the case of simple random sampling were discussed. In this edition, we also include remedies for extreme values including the result of Zelterman (1993) and the use of the  $m$  out of  $n$  bootstrap.

Bootstrap diagnostics are also discussed in Chapter 9. Efron’s jackknife-after-bootstrap is discussed because it is the first tool devised to help identify

whether or not a nonparametric bootstrap will work in a given application. The work from Efron (1992c) is described in Section 9.7.

Chapter 9 differs from the other chapters in that it goes into some of the technical probability details that the practitioner lacking this background may choose to skip. The practitioner may not need 1992c to understand exactly why these cases fail but should have a general awareness of the cases where the ordinary bootstrap fails and whether or not remedies have been found.

Each chapter (except Chapter 6) has a historical notes section. This section is intended as a guide to the literature related to the chapter and puts the results into their chronological order of development. I found that this was a nice feature in several earlier bootstrap books, including Hall (1992a), Efron and Tibshirani (1993), and Davison and Hinkley (1997). Although related references are cited throughout the text, the historical notes are intended to provide a perspective regarding when the techniques were originally proposed and how the key developments followed chronologically.

One notable change in the second edition is the increased description of techniques, particularly in Chapters 8 and 9.

## 1.2. INTRODUCTION

Two of the most important problems in applied statistics are the determination of an estimator for a particular parameter of interest and the evaluation of the accuracy of that estimator through estimates of the standard error of the estimator and the determination of confidence intervals for the parameter. Efron, when introducing his version of the “bootstrap” (Efron, 1979a), was particularly motivated by these two problems. Most important was the estimation of the standard error of the parameter estimator, particularly when the estimator was complex and standard approximations such as the delta methods were either not appropriate or too inaccurate.

Because of the bootstrap’s generality, it has been applied to a much wider class of problems than just the estimation of standard errors and confidence intervals. Applications include error rate estimation in discriminant analysis, subset selection in regression, logistic regression, and classification problems, cluster analysis, kriging (i.e., a form of spatial modeling), nonlinear regression, time series analysis, complex surveys,  $p$ -value adjustment in multiple testing problems, and survival and reliability analysis.

It has been applied in various disciplines including psychology, geology, econometrics, biology, engineering, chemistry, and accounting. It is our purpose to describe some of these applications in detail for the practitioner in order to exemplify its usefulness and illustrate its limitations. In some cases the bootstrap will offer a solution that may not be very good but may still be used for lack of an alternative approach. Since the publication of the first edition of this text, research has emphasized applications and has added to the long list of applications including particular applications in the pharma-

ceutical industry. In addition, modifications to the bootstrap have been devised that overcome some of the limitations that had been identified.

Before providing a formal definition of the bootstrap, here is an informal description of how it works. In its most general form, we have a sample of size  $n$  and we want to estimate a parameter or determine the standard error or a confidence interval for the parameter or even test a hypothesis about the parameter. If we do not make any parametric assumptions, we may find this difficult to do. The bootstrap provides a way to do this.

We look at the sample and consider the empirical distribution. The empirical distribution is the probability distribution that has probability  $1/n$  assigned to each sample value. The bootstrap idea is simply to replace the unknown population distribution with the known empirical distribution.

Properties of the estimator such as its standard error are then determined based on the empirical distribution. Sometimes these properties can be determined analytically, but more often they are approximated by Monte Carlo methods (i.e., we sample with replacement from the empirical distribution).

Now here is a more formal definition. Efron's bootstrap is defined as follows: Given a sample of  $n$  independent identically distributed random vectors  $X_1, X_2, \dots, X_n$  and a real-valued estimator  $(X_1, X_2, \dots, X_n)$  (denoted by  $\hat{\theta}$ ) of the parameter  $\theta$ , a procedure to assess the accuracy of  $\hat{\theta}$  is defined in terms of the empirical distribution function  $F_n$ . This empirical distribution function assigns probability mass  $1/n$  to each observed value of the random vectors  $X_i$  for  $i = 1, 2, \dots, n$ .

The empirical distribution function is the maximum likelihood estimator of the distribution for the observations when no parametric assumptions are made. The bootstrap distribution for  $\hat{\theta} - \theta$  is the distribution obtained by generating  $\hat{\theta}$ 's by sampling independently with replacement from the empirical distribution  $F_n$ . The bootstrap estimate of the standard error of  $\hat{\theta}$  is then the standard deviation of the bootstrap distribution for  $\hat{\theta} - \theta$ .

It should be noted here that almost any parameter of the bootstrap distribution can be used as a "bootstrap" estimate of the corresponding population parameter. We could consider the skewness, the kurtosis, the median, or the 95th percentile of the bootstrap distribution for  $\hat{\theta}$ .

Practical application of the technique usually requires the generation of bootstrap samples or resamples (i.e., samples obtained by independently sampling with replacement from the empirical distribution). From the bootstrap sampling, a Monte Carlo approximation of the bootstrap estimate is obtained. The procedure is straightforward.

1. Generate a sample with replacement from the empirical distribution (a bootstrap sample),
2. Compute \* the value of  $\hat{\theta}$  obtained by using the bootstrap sample in place of the original sample,
3. Repeat steps 1 and 2  $k$  times.

For standard error estimation,  $k$  is recommended to be at least 100. This recommendation can be attributed to the article Efron (1987). It has recently been challenged in a paper by Booth and Sarkar (1998). Further discussion on this recommendation can be found in Chapter 7.

By replicating steps 1 and 2  $k$  times, we obtain a Monte Carlo approximation to the distribution of  $\theta^*$ . The standard deviation of this Monte Carlo distribution of  $\theta^*$  is the Monte Carlo approximation to the bootstrap estimate of the standard error for  $\hat{\theta}$ . Often this estimate is simply referred to as the bootstrap estimate, and for  $k$  very large (e.g., 500) there is very little difference between the bootstrap estimator and this Monte Carlo approximation.

What we would like to know for inference is the distribution of  $\hat{\theta} - \theta$ . What we have is a Monte Carlo approximation to the distribution of  $\theta^* - \hat{\theta}$ . The key idea of the bootstrap is that for  $n$  sufficiently large, we expect the two distributions to be nearly the same.

In a few cases, we are able to compute the bootstrap estimator directly without the Monte Carlo approximation. For example, in the case of the estimator being the mean of the distribution of a real-valued random variable, Efron (1982a, p. 2) states that the bootstrap estimate of the standard error of is  $\hat{\sigma}_{\text{BOOT}} = [(n-1)/n]^{1/2} \hat{\sigma}$ , where  $\hat{\sigma}$  is defined as

$$\hat{\sigma} = \left[ \frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2},$$

where  $x_i$  is the value of the  $i$ th observation and  $\bar{x}$  is the mean of the sample. As a second example, consider the case of testing the hypothesis of equality of distributions for censored matched pairs (i.e., observations whose values may be truncated). The bootstrap test applied to paired differences is equivalent to the sign test and the distribution under the null hypothesis is binomial with  $p = 1/2$ . So no bootstrap sampling is required to determine the critical region for the test.

The bootstrap is often referred to as a computer-intensive method. It gets this label because in most practical problems where it is deemed to be useful the estimation is complex and bootstrap samples are required. In the case of confidence interval estimation and hypothesis testing problems, this may mean at least 1000 bootstrap replications (i.e.,  $k = 1000$ ). In Section 7.1, we address the important practical issue of what value to use for  $k$ .

Methods for reducing the computer time by more efficient Monte Carlo sampling are discussed in Section 7.2. The examples above illustrate that there are cases for which the bootstrap is not computer-intensive at all!

Another point worth emphasizing here is that the bootstrap samples differ from the original sample because some of the observations will be repeated once, twice, or more in a bootstrap sample. There will also be some observations that will not appear at all in a particular bootstrap sample. Consequently, the values for  $\theta^*$  will vary from one bootstrap sample to the next.