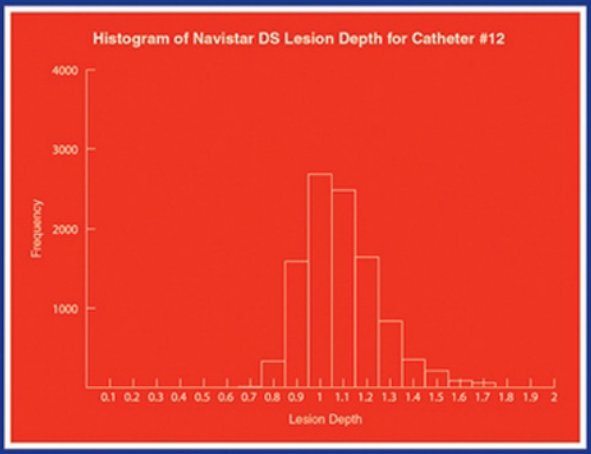
WILEY

Bootstrap Methods

A Guide for Practitioners and Researchers

Second Edition



Michael R. Chernick

WILEY SERIES IN PROBABILITY AND STATISTICS

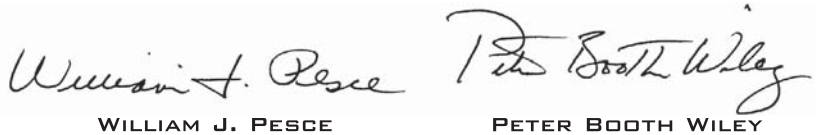
Bootstrap Methods

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

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

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

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.

CHAPTER 1

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

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

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 skimped 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 ($k = 1, 2, 3, \ldots, n$) 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, Dmitris 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 $X_1, X_2, \ldots, X_n$ and a real-valued estimator $(X_1, X_2, \ldots, X_n)$ (denoted by $\hat{\theta}$) of the parameter , 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 $i = 1, 2, \ldots, n$.

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 θ^*. The standard deviation of this Monte Carlo distribution of θ^* 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

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

where x_i is the value of the ith 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 θ^* will vary from one bootstrap sample to the next.

The actual probability that a particular X_i will appear j times in a bootstrap sample for $j = 0, 1, 2, \ldots, n$, can be determined using the multinomial distribution or alternatively by using classical occupancy theory. For the latter approach see (Chernick and Murthy, 1985). Efron (1983) calls these probabilities the repetition rates and discusses them in motivating the use of the .632 estimator (a particular bootstrap type estimator) for classification error rate estimation. A general account of the classical occupancy problem can be found in Johnson and Kotz (1977).

The basic idea behind the bootstrap is the variability of θ^* (based on F_n) around $\hat{\theta}$ will be similar to (or mimic) the variability of $\hat{\theta}$ (based on the true population distribution F) around the true parameter value, θ. There is good reason to believe that this will be true for large sample sizes, since as n gets larger and larger, F_n comes closer and closer to F and so sampling with replacements from F_n is almost like random sampling from F.

The strong law of large numbers for independent identically distributed random variables implies that with probability one, F_n converges to F pointwise [see Chung (1974, pp. 131–132) for details]. Strong laws pertaining to the bootstrap can be found in Athreya (1983). A stronger result, the Glivenko–Cantelli theorem [see Chung (1974, p. 133)], asserts that the empirical distribution converges uniformly with probability 1 to the distribution F when the observations are independent and identically distributed. Although not stated explicitly in the early bootstrap literature, this fundamental theoretical result lends credence to the bootstrap approach. The theorem was extended in Tucker (1959) to the case of a random sequence from a strictly stationary stochastic process.

In addition to the Glivenko–Cantelli theorem, the validity of the bootstrap requires that the estimator (a functional of the empirical distribution function) converge to the "true parameter value" (i.e., the functional for the "true" population distribution). A functional is simply a mapping that assigns a real value to a function. Most commonly used parameters of distribution functions can be expressed as functionals of the distribution, including the mean, the variance, the skewness, and the kurtosis.

Interestingly, sample estimates such as the sample mean can be expressed as the same functional applied to the empirical distribution. For more discussion of this see Chernick (1982), who deal with a form of a functional derivative called an influence function. The concept of an influence function was first introduced by Hampel (1974) as a method for comparing robust estimators.

Influence functions have had uses in robust statistical methods and in the detection of outlying observations in data sets. Formal treatment of statistical functionals can be found in Fernholtz (1983). There are also connections for the influence function with the jackknife and the bootstrap as shown by Efron (1982a).

Convergence of the bootstrap estimate to the appropriate limit (consistency) requires some sort of smoothness condition on the functional corresponding to the estimator. In particular, conditions given in Hall (1992a)

employ asymptotic normality for the functional and further allow for the existence of an Edgeworth expansion for its distribution function. So there is more needed. For independent and identically distributed observations we require (1) the convergence of F_n to F (this is satisfied by virtue of the Glivenko–Cantelli theorem), (2) an estimate that is the corresponding functional of F_n as the parameter is of F (satisfied for means, standard deviations, variances, medians, and other sample quantiles of the distribution), and (3) a smoothness condition on the functional. Some of the consistency proofs also make use of the well-known Berry–Esseen theorem [see Lahiri (2003a, pp. 21–22, Theorem 2.1) for the sample mean]. When the bootstrap fails (i.e., bootstrap estimates are inconsistent), it is often because the smoothness condition is not satisfied (e.g., extreme order statistics such as the minimum or maximum of the sample).

These Edgeworth expansions along with the Cornish–Fisher expansions not only can be used to assure the consistency of the bootstrap, but they also provide asymptotic rates of convergence. Examples where the bootstrap fails asymptotically, due to a lack of smoothness of the functional, are given in Chapter 9.

Also, the original bootstrap idea applies to independent identically distributed observations and is guaranteed to work only in large samples. Using the Monte Carlo approximation, bootstrapping can be applied to many practical problems such as parameter estimation in time series, regression, and analysis of variance problems, and even to problems involving small samples.

For some of these problems, we may be on shaky ground, particularly when small sample sizes are involved. Nevertheless, through the extensive research that took place in the 1980s and 1990s, it was discovered that the bootstrap sometimes works better than conventional approaches even in small samples (e.g., the case of error rate estimation for linear discriminant functions to be discussed in Section 2.1.2).

There is also a strong temptation to apply the bootstrap to a number of complex statistical problems where we cannot resort to classical theory to resort to. At least for some of these problems, we recommend that the practitioner try the bootstrap. Only for cases where there is theoretical evidence that the bootstrap leads us astray would we advise against its use.

The determination of variability in subset selection for regression, logistic regression, and its use in discriminant analysis problems provide examples of such complex problems. Another example is the determination of the variability of spatial contours based on the method of kriging. The bootstrap and alternatives in spatial problems are treated in Cressie (1991). Other books that cover spatial data problems are Mardia, Kent, and Bibby (1979) and Hall (1988c). Tibshirani (1992) provides some examples of the usefulness of the bootstrap in complex problems.

Diaconis and Efron (1983) demonstrate, with just five bootstrap sample contour maps, the value of the bootstrap approach in uncovering the vari-

ability in the contours. These problems that can be addressed by the bootstrap approach are discussed in more detail in Chapter 8.

1.3. WIDE RANGE OF APPLICATIONS

As mentioned at the end of the last section, there is a great deal of temptation to apply the bootstrap in a wide number of settings. In the regression case, for example, we may treat the vector including the dependent variable and the explanatory variable as independent random vectors, or alternatively we may compute residuals and bootstrap them. These are two distinct approaches to bootstrapping in regression problems which will be discussed in detail in Chapter 5.

In the case of estimating the error rate of a linear discriminant function, Efron showed in Efron (1982a, pp. 49–58) and Efron (1983) that the bootstrap could be used to (1) estimate the bias of the "apparent error rate" estimate (a naïve estimate of error rate that is also referred to as the resubstitution estimate) and (2) produce an improved error rate estimate by adjusting for the bias.

The most attractive feature of the bootstrap and the permutation tests described in Good (1994) is the freedom they provide from restrictive parametric assumptions and simplified models. There is no need to force Gaussian or other parametric distributional assumptions on the data.

In many problems, the data may be skewed or have a heavy-tailed distribution or may even be multimodal. The model does not need to be simplified to some "linear" approximation, and the estimator itself can be complicated.

We do not require an analytic expression for the estimator. The bootstrap Monte Carlo approximation can be applied as long as there is a computational method for deriving the estimator. That means that we can numerical integrate using iterative schemes to calculate the estimator. The bootstrap doesn't care. The only price we pay for such complications is in the time and cost for the computer usage (which is becoming cheaper and faster).

Another feature that makes the bootstrap approach attractive is its simplicity. We can formulate bootstrap simulations for almost any conceivable problem. Once we program the computer to carry out the bootstrap replications, we let the computer do all the work. A danger to this approach is that a practitioner might bootstrap at will, without consulting a statistician (or considering the statistical implications) and without giving careful thought to the problem.

This book will aid the practitioner in the proper use of the bootstrap by acquainting him with its advantages and limitations, lending theoretical support where available and Monte Carlo results where the theory is not yet available. Theoretical counterexamples to the consistency of bootstrap estimates also provide guidelines to its limitations and warn the practitioner when not to

apply the bootstrap. Some simulation studies also provide such negative results.

However, over the past 9 years, modifications to the basic or naïve bootstrap that fails due to inconsistency have been constructed to be consistent. One notable approach to be covered in Chapter 9 is the *m*-out-of-*n* bootstrap. Instead of sampling *n* times with replacement from the empirical distribution where *n* is the original sample size, the *m*-out-of-*n* bootstrap samples *m* times with replacement from the empirical distribution where *m* is chosen to be less than *n*. In the asymptotic theory both *m* and *n* tend to infinity but *m* increases at a slower rate. The rate to choose depends on the application.

I believe, as do many others now, that many simulation studies indicate that the bootstrap can safely be applied to a large number of problems even where strong theoretical justification does not yet exist. For many problems where realistic assumptions make other statistical approaches impossible or at least intractable, the bootstrap at least provides a solution even if it is not a very good one. For some people in certain situations, even a poor solution is better than no solution.

Another problem that creates difficulties for the scientist and engineer is that of missing data. In designing an experiment or a survey, we may strive for balance in the design and choose specific samples sizes in order to make the planned inferences from the data. The correct inference can be made only if we observe the complete data set.

Unfortunately, in the real world, the cost of experimentation, faulty measurement, or lack of response from those selected for the survey may lead to incomplete and possibly unbalanced designs. Milliken and Johnson (1984) refer to such problem data as messy data.

In Milliken and Johnson (1984, 1989) they provide ways to analyze messy data. When data are missing or censored, bootstrapping provides another approach for dealing with the messy data (see Section 8.4 for more details on censored data, and see Section 8.7 for an application to missing data).

The bootstrap alerts the practitioner to variability in his data, of which he or she may not be aware. In regression, logistic regression, or discriminant analysis, stepwise subset selection is a commonly used method available in most statistical computer packages. The computer does not tell the user how arbitrary the final selection actually is. When a large number of variables or features are included and many are correlated or redundant, there can be a great deal of variability to the selection. The bootstrap samples enable the user to see how the chosen variables or features change from bootstrap sample to bootstrap sample and provide some insight as to which variables or features are really important and which ones are correlated and easily substituted for by others. This is particularly well illustrated by the logistic regression problem studied in Gong (1986). This problem is discussed in detail in Section 8.2.

In the case of kriging, spatial contours of features such as pollution concentration are generated based on data at monitoring stations. The method is a

form of interpolation between the stations based on certain statistical spatial modeling assumptions. However, the contour maps themselves do not provide the practitioner with an understanding of the variability of these estimates. Kriging plots for different bootstrap samples provide the practitioner with a graphical display of this variability and at least warn him of variability in the data and analytic results. Diaconis and Efron (1983) make this point convincingly, and I will demonstrate this application in Section 8.1. The practical value of this cannot be underestimated!

Babu and Feigelson (1996) discuss applications in astronomy. They devote a whole chapter (Chapter 5, pp. 93–103) to resampling methods, emphasizing the importance of the bootstrap.

In clinical trials, sample sizes are determined based on achieving a certain power for a statistical hypothesis of efficacy of the treatment. In Section 3.3, I show an example of a clinical trial for a pacemaker lead (Pacesetter's Tendril DX model). In this trial, the sample sizes for the treatment and control leads were chosen to provide an 80% chance of detecting a clinically significant improvement (decrease of 0.5 volts) in the average capture threshold at the three-month follow-up for the experimental Tendril DX lead (model 1388T) compared to the respective control lead (Tendril model 1188T) when applying a one-sided significance test at the 5% significance level. This was based on the standard normal distribution theory. In the study, nonparametric methods were also considered. Bootstrap confidence intervals based on Efron's percentile method were used to do the hypothesis test without needing parametric assumptions. The Wilcoxon rank sum test was another nonparametric procedure that was used to test for a statistically significant change in capture threshold.

A similar study for a passive fixation lead, the Passive Plus DX lead, was conducted to get FDA approval for the steroid eluting version of this type of lead. In addition to comparing the investigational (steroid eluting) lead with the non-steroid control lead, using both the bootstrap (percentile method) and Wilcoxon rank sum tests, I also tried the bootstrap percentile t confidence intervals for the test. This method theoretically can give a more accurate confidence interval. The results were very similar and conclusive at showing the efficacy of the steroid lead. The percentile *t* method of confidence interval estimation is described in Section 3.1.5.

However, the statistical conclusion for such a trial is based on a single test at the three-month follow-up after all 99 experimental and 33 control leads have been implanted, and the patients had threshold tests at the three-month follow-up.

In the practice of clinical trials, the investigators do not want to wait for all the patients to reach their three-month follow-up before doing the analysis. Consequently, it is quite common to do interim analyses at some point or points in the trial (it could be one in the middle of the trial or two at the one-third and two-thirds points in the trial). Also, separate analyses are sometimes done on subsets of the population. Furthermore, sometimes separate analyses

are done on subsets of the population. These examples are all situations where multiple testing is involved. Multiple testing requires specific techniques for controlling the type I error rate (in this context the so-called family-wise error rate is the error rate that is controlled. Equivalent to controlling the family-wise type I error rate the *p*-values for the individual tests can be adjusted. Probability bounds such as the Bonferroni can be used to give conservative estimates of the *p*-value or simultaneous inference methods can be used [see Miller (1981b) for a thorough treatment of this subject].

An alternative approach would be to estimate the *p*-value adjustment by bootstrapping. This idea has been exploited by Westfall and Young and is described in detail in Westfall and Young (1993). We will attempt to convey the key concepts. The application of bootstrap *p*-value adjustment to the Passive Plus DX clinical trial data is covered in Section 8.5. Consult Miller (1981b), Hsu (1996), and/or Westfall and Young (1993) for more details on multiple testing, *p*-value adjustment, and multiple comparisons.

In concluding this section, we wish to emphasize that the bootstrap is not a panacea. There are certainly practical problems where classical parametric methods are reasonable and provide either more efficient estimates or more powerful hypothesis tests. Even for some parametric problems, the parametric bootstrap, as discussed by Davison and Hinkley (1997, p. 3) and illustrated by them on pages 148 and 149, can be useful.

What the bootstrap does do is free the scientist from restrictive modeling and distributional assumptions by using the power of the computer to replace difficult analysis. In an age when computers are becoming more and more powerful, inexpensive, fast, and easy to use, the future looks bright for additional use of these so-called computer-intensive statistical methods, as we have seen over the past decade.

1.4. HISTORICAL NOTES

It should be pointed out that bootstrap research began in the late 1970s, although many key related developments can be traced back to earlier times. Most of the important theoretical development; took place in the1980s after Efron (1979a). The first proofs of the consistency of the bootstrap estimate of the sample mean came in 1981 with the papers of Singh (1981) and Bickel and Freedman (1981).

Regarding this seminal paper by Efron (1979a), Davison and Hinkley (1997) write "The publication in 1979 of Bradley Efron's first article on bootstrap methods was a major event in Statistics, at once synthesizing some of the earlier resampling ideas and establishing a new framework for simulation-based statistical analysis. The idea of replacing complicated and often inaccurate approximations to biases, variances, and other measures of uncertainty by computer simulations caught the imagination of both theoretical researchers and users of statistical methods."

As mentioned earlier in this chapter, a number of related techniques are often referred to as resampling techniques. These other resampling techniques predate Efron's bootstrap. Among these are the jackknife, cross-validation, random subsampling, and the permutation test procedures described in Good (1994), Edgington (1980, 1987, 1995), and Manly (1991, 1997).

Makinodan, Albright, Peter, Good, and Heidrick (1976) apply permutation tests to study the effect of age in mice on the mediation of immune response. Due to the fact that an entire factor was missing, the model and the permutation test provides a clever way to deal with imbalance in the data. A detailed description is given in Good (1994, pp. 58–59).

Efron himself points to some of the early work of R. A. Fisher (in the 1920s) on maximum likelihood estimation as the inspiration for many of the basic ideas. The jackknife was introduced by Quenouille (1949) and popularized by Tukey (1958), and Miller (1974) provides an excellent review of the jackknife methods. Extensive coverage of the jackknife can be found in the book by Gray and Schucany (1972).

Bickel and Freedman (1981) and Singh (1981) presented the first results demonstrating the consistency of the bootstrap undercertain mathematical conditions. Bickel and Freedman (1981) also provide a counterexample for consistency of the nonparametric bootstrap, and this is also illustrated by Schervish (1995, p. 330, Example 5.80). Gine and Zinn (1989) provide necessary conditions for the consistency of the bootstrap for the mean.

Athreya (1987a,b), Knight (1989), and Angus (1993) all provide examples where the bootstrap failed to be consistent due to its inability to meet certain necessary mathematical conditions. Hall, Hardle, and Simar (1993) showed that estimators for bootstrap distributions can also be inconsistent.

The general subject of empirical processes is related to the bootstrap and can be used as a tool to demonstrate consistency (see Csorgo, 1983; Shorack and Wellner, 1986; van der Vaart and Wellner, 1996). Fernholtz (1983) provides the mathematical theory of statistical functionals and functional derivatives (such as influence functions) that relate to bootstrap theory.

Quantile estimation via bootstrapping appears in Helmers, Janssen, and Veraverbeke (1992) and in Falk and Kaufmann (1991). Csorgo and Mason (1989) bootstrap the empirical distribution and Tu (1992) uses jackknife pseudovalues to approximate the distribution of a general standardized functional statistic.

Subsampling methods began with Hartigan (1969, 1971, 1975) and McCarthy (1969). These papers are discussed briefly in the development of bootstrap confidence intervals in Chapter 3. A more recent account is given by Babu (1992).

Young and Daniels (1990) discuss the bias that is introduced in Efron's nonparametric bootstrap by the use of the empirical distribution as a substitute for the true unknown distribution.

Diaconis and Holmes (1994) show how to avoid the Monte Carlo approximation to the bootstrap by cleverly enumerating all possible bootstrap samples using what are called Gray codes.

The term bootstrap has been used in other similar contexts which predate Efron's work, but these methods are not the same and some confusion occurs. When I gave a presentation on the bootstrap at the Aerospace Corporation in 1983 a colleague, Dr. Ira Weiss, mentioned that he used the bootstrap in 1970 long before Efron coined the term. After looking at Ira's paper, I realized that it was a different procedure with a similar idea.

Apparently, control theorists came up with a procedure for applying Kalman filtering with an unknown noise covariance which they also named the bootstrap. Like Efron, they were probably thinking of the old adage "picking yourself up by your own bootstraps" (as was attributed to the fictional Baron von Munchausen as a trick for climbing out from the bottom of a lake) when they chose the term to apply to an estimation procedure that avoids a priori assumptions and uses only the data at hand. A survey and comparison of procedures for dealing with the problem of unknown noise covariance including this other bootstrap technique is given in Weiss (1970). The term bootstrap has also been used in totally different contexts by computer scientists.

An entry on bootstrapping in the *Encyclopedia of Statistical Science* (1981, Volume 1, p. 301) is provided by the editors and is very brief. In 1981 when that volume was published, the true value of bootstrapping was not fully appreciated. The editors subsequently remedied this with an article in the supplemental volume.

The point, however, is that the original entry cited only three references. The first, Efron's *SIAM Review* article (Efron, 1979b), was one of the first published works describing Efron's bootstrap. The second article from *Technometrics* by Fuchs (1978) does not appear to deal with the bootstrap at all! The third article by LaMotte (1978) and also in *Technometrics* does refer to a bootstrap but does not mention any of Efron's ideas and appears to be discussing a different bootstrap.

Because of these other bootstraps, we have tried to refer to the bootstrap as Efron's bootstrap; a few others have done the same, but it has not caught on. In the statistical literature, reference to the bootstrap will almost always mean Efron's bootstrap or some derivative of it. In the engineering literature an ambiguity may exist and we really need to look at the description of the procedure in detail to determine precisely what the author means.

The term bootstrap has also commonly appeared in the computer science literature, and I understand that mathematicians use the term to describe certain types of numerical solutions to partial differential equations. Still it is my experience that if I search for articles in mathematical or statistical indices using the keyword "bootstrap," I would find that the majority of the articles referred to Efron's bootstrap or a variant of it. I wrote the preceding statement back in 1999 when the first edition was published. Now in 2007, I formed the basis for the second bibliography of the text by searching the Current Index

to Statistics (CIS) for the years 1999 to 2007 with only the keyword "bootstrap" required to appear in the title or the list of key words. Of the large number of articles and books that I found from this search, all of the references were referring to Efron's bootstrap or a method derived from the original idea of Efron. The term "boofstrap" is used these days as a noun or a verb.

However, I have no similar experience with the computer science literature or the engineering literature. But Efron's bootstrap now has a presence in these two fields as well. In computer science there have been many meetings on the interface between computer science and statistics, and much of the common ground involves computer-intensive methods such as the bootstrap. Because of the rapid growth of bootstrap application in a variety of industries, the "statistical" bootstrap now appears in some of the physics and engineering journals including the IEEE journals. In fact the article I include in Chapter 4, an application of nonlinear regression to a quasi-optical experiment, I coauthored with three engineers and the article appeared in the *IEEE Transactions on Microwave Theory and Techniques*.

Efron (1983) compared several variations to the bootstrap estimate. He considered simulation of Gaussian distributions for the two-class problem (with equal covariances for the classes) and small sample sizes (e.g., a total of, say, 14–20 training samples split equally among the two populations). For linear discriminant functions, he showed that the bootstrap and in particular the .632 estimator are superior to the commonly used leave-one-out estimate (also called cross-validation by Efron). Subsequent simulation studies will be summarized in Section 2.1.2 along with guidelines for the use of some of the bootstrap estimates.

There have since been a number of interesting simulation studies that show the value of certain bootstrap variants when the training sample size is small (particularly the estimator referred to as the .632 estimate). In a series of simulations studies, Chernick, Murthy, and Nealy (1985, 1986, 1988a,b) confirmed the results in Efron (1983). They also showed that the .632 estimator was superior when the populations were not Gaussian but had finite first moments. In the case of Cauchy distributions and other heavy-tailed distributions from the Pearson VII family of distributions which do not have finite first moments, they showed that other bootstrap approaches were better than the .632 estimator.

Other related simulation studies include Chatterjee and Chatterjee (1983), McLachlan (1980), Snapinn and Knoke (1984, 1985a,b, 1988), Jain, Dubes, and Chen (1987) and Efron and Tibshirani (1997a). We summarize the results of these studies and provide guidelines to the use of the bootstrap procedures for linear and quadratic discriminant functions in Section 2.1.2. McLachlan (1992) also gives a good summary treatment to some of this literature. Additional theoretical results can be found in Davison and Hall (1992). Hand (1986) is another good survey article on error rate estimation. The 632+ estimator proposed by Efron and Tibshirani (1997a) was applied to an ecological

problem by Furlanello, Merler, Chemini, and Rizzoli (1998). Ueda and Nakano (1995) apply the bootstrap and cross-validation to error rate estimation for neural network-type classifiers. Hand (1981, p. 189; 1982, pp. 178–179) discusses the bootstrap approach to estimating the error rates in discriminant analysis.

In the late 1980s and the 1990s, a number of books appeared that covered some aspect of bootstrapping at least partially. Noreen's book (Noreen, 1989) deals with the bootstrap in very elementary ways for hypothesis testing only.

There are now several survey articles on bootstrapping in general, including Babu and Rao (1993), Young (1994), Stine (1992), Efron (1982b), Efron and LePage (1992), Efron and Tibshirani (1985, 1986, 1996a, 1997b), Hall (1994), Manly (1993), Gonzalez-Manteiga, Prada-Sanchez, and Romo (1993), Politis (1998), and Hinkley (1984, 1988). Overviews on the bootstrap or special aspects of bootstrapping include Beran (1984b), Leger, Politis, and Romano (1992), Pollack, Simon, Bruce, Borenstein, and Lieberman (1994), and Fiellin and Feinstein (1998) on the bootstrap in general; Babu and Bose (1989), DiCiccio and Efron (1996), and DiCiccio and Romano (1988, 1990) on confidence intervals; Efron (1988b) on regression; Falk (1992a) on quantile estimation; and DeAngelis and Young (1992) on smoothing. Lanyon (1987) reviews the jackknife and bootstrap for applications to ornithology. Efron (1988c) gives a general discussion of the value of bootstrap confidence intervals aimed at an audience of psychologists.

The latest edition of *Kendall's Advanced Theory of Statistics*, Volume I, deals with the bootstrap as a tool for estimating standard errors in Chapter 10 [see Stuart and Ord (1993, pp. 365–368)].

The use of the bootstrap to compute standard errors for estimates and to obtain confidence intervals for multilevel linear models is given in Goldstein (1995, pp. 60–63). Waclawiw and Liang (1994) give an example of parametric bootstrapping using generalized estimating equations. Other works involving the bootstrap and jackknife in estimating equation models include Lele (1991a,b).

Lehmann and Casella (1998) mention the bootstrap as a tool in reducing the bias of an estimator (p. 144) and in the attainment of higher order efficiency (p. 519). Lehmann (1999, Section 6.5, pp. 420–435) presents some details on the asymptotic properties of the bootstrap.

In the context of generalized least-squares estimation of regression parameters Carroll and Ruppert (1988, pp. 26–28) describe the use of the bootstrap to get confidence intervals. In a brief discussion, Nelson (1990) mentions the bootstrap as a potential tool in regression models with right censoring of data for application to accelerated lifetime testing. Srivastava and Singh (1989) deal with the application of bootstrap in multiplicative models. Bickel and Ren (1996) employ an m-out-of-n bootstrap for goodness of fit tests with doubly censored data.

McLachlan and Basford (1988) discuss the bootstrap in a number of places as an approach for determining the number of distributions or modes in a

mixture model. Another excellent text on mixture models is Titterington, Smith, and Makov (1985). Efron and Tibshirani (1996b) take a novel approach to bootstrapping that can be applied to the determination of the number of modes in a density function and the number of variables in a model. In addition to determining the number of modes, Romano (1988c) uses the bootstrap to determine the location of a mode.

Linhart and Zucchini (1986, pp. 22–23) describe how the bootstrap can be used for model selection. Thompson (1989, pp. 42–43) mentions the use of bootstrap techniques for estimating parameters in growth models (i.e., a nonlinear regression problem). McDonald (1982) shows how smoothed or ordinary bootstrap samples can be drawn to obtain regression estimates.

Rubin (1987, pp. 44–46) discusses his "Bayesian" bootstrap for problems of imputation. The original paper on the Bayesian bootstrap is Rubin (1981). Banks (1988) provides a modification to the Bayesian bootstrap. Other papers involving the Bayesian bootstrap are Lo (1987, 1988, 1993a) and Weng (1989). Geisser (1993) discusses the bootstrap with respect to predictive distributions (another Bayesian concept). Ghosh and Meeden (1997, pp. 140–149) discuss applications of the Bayesian bootstrap to finite population sampling. The Bayesian bootstrap is often applied to imputation problems. Rubin (1996) is a survey article detailing the history of multiple imputation. At the time of the article the method of multiple imputation had been studied for more than 18 years.

Rey (1983) devotes Chapter 5 of his monograph to the bootstrap. He is using it in the context of robust estimation. His discussion is particularly interesting because he mentions both the pros and the cons and is critical of some of the early claims made for the bootstrap [particularly in Diaconis and Efron (1983)].

Staudte and Sheather (1990) deal with the bootstrap as an approach to estimating standard errors of estimates. They are particularly interested in the standard errors of robust estimators. Although they do deal with hypothesis testing, they do not use the bootstrap for any hypothesis testing problems. Their book includes a computer disk that has Minitab macros for bootstrapping in it. Minitab computer code for these macros is presented in Appendix D of their book.

Barnett and Lewis (1995) discuss the bootstrap as it relates to checking modeling assumptions in the face of outliers. Agresti (1990) discusses the bootstrap as it can be applied to categorical data.

McLachlan and Krishnan (1997) discuss the bootstrap in the context of robust estimation of a covariance matrix. Beran and Srivastava (1985) provide bootstrap tests for functions of a covariance matrix. Other papers covering the theory of the bootstrap as it relates to robust estimators are Babu and Singh (1984b) and Arcones and Gine (1992). Lahiri (1992a) does bootstrapping of *M*-estimators (a type of robust location estimator).

The text by van der Vaart and Wellner (1996) is devoted to weak convergence and empirical processes. Empirical process theory can be applied to

obtain important results in bootstrapping, and van der Vaart and Wellner illustrate this in Section 3.6 of their book (14 pages devoted to the subject of bootstrapping, pp. 345–359).

Hall (1992a) considers functionals that admit Edgeworth expansions. Edgeworth expansions provide insight into the accuracy of bootstrap confidence intervals, the value of bootstrap hypothesis tests, and use of the bootstrap in parametric regression. It also provides guidance to the practitioner regarding the variants of the bootstrap and the Monte Carlo approximations. Some articles relating Edgeworth expansions to applications of the bootstrap include Abramovitch and Singh (1985), Bhattacharya and Qumsiyeh (1989), Babu and Singh (1989), and Bai and Rao (1991, 1992).

Chambers and Hastie (1991) discuss applications of statistical models through the use of the *S* language. They discuss the bootstrap in various places.

Gifi (1990) applies the bootstrap to multivariate problems. Other uses of the bootstrap in branches of multivariate analysis are are documented Diaconis and Efron (1983), who apply the bootstrap to principal component analysis, and Greenacre (1984), who covers the use of bootstrapping in correspondence analysis.

One of the classic texts on multivariate analysis is Anderson (1959), which was the first to provide extensive coverage of the theory based on the multivariate normal model. In the second edition of the text, Anderson (1984), he introduces some bootstrap applications. Flury (1997) provides another recent account of multivariate analysis. Flury (1988) is a text devoted to the principal components technique and so is Jolliffe (1986). Seber (1984), Gnandesikan (1977, 1997), Hawkins (1982), and Mardia, Kent, and Bibby (1979) all deal with the subject of multivariate analysis and multivariate data.

Scott (1992, pp. 257–260) discusses the bootstrap as a tool in estimating standard errors and confidence intervals in the context of multivariate density estimation. Other articles where the bootstrap appears as a density estimation tool are Faraway and Jhun (1990), Falk (1992b), and Taylor and Thompson (1992).

Applications in survival analysis include Burr (1994), Hsieh (1992), LeBlanc and Crowley (1993) and Gross and Lai (1996a).

An application of the double bootstrap appears in McCullough and Vinod (1998). Application to the estimation of correlation coefficients can be found in Lunneborg (1985) and Young (1988a).

General discussion of bootstrapping related to nonparametric procedures include Romano (1988a), Romano (1989b), and Simonoff (1986), where goodness of fit of distributions in sparse multinomial data problems is addressed using the bootstrap. Tu, Burdick, and Mitchell (1992) apply bootstrap resampling to nonparametric rank estimation.

Hahn and Meeker (1991) briefly discuss bootstrap confidence intervals. Frangos and Schucany (1990) discuss the technical aspects of estimating the acceleration constant for Efron's BC_a confidence interval method. Bickel and

Krieger (1989) use the bootstrap to attain confidence bands for a distribution function, and Wang and Wahba (1995) get bootstrap confidence bands for smoothing splines and compare them to bands constructed using Bayesian methods.

Bailey (1992) provides a form of bootstrapping for order statistics and other random variables whose distributions can be represented as convolutions of other distributions. By substituting the empirical distributions for the distributions in the convolution, a "bootstrap" distribution for the random variable is derived.

Beran (1982) compares the bootstrap with various competitive methods in estimating sampling distributions. Bau (1984) does bootstrapping for statistics involving linear combinations. Parr (1983) is an early reference comparing the bootstrap, the jackknife, and the delta method in the context of bias and variance estimation. Hall (1988d) deals with the rate of convergence for bootstrap approximations.

Applications to directional data include Fisher and Hall (1989) and Ducharme, Jhun, Romano, and Troung (1985). Applications to finite population sampling include Chao and Lo (1985), Booth, Butler, and Hall (1994), Kuk (1987, 1989), and Sitter (1992b).

Applications have appeared in a variety of disciplines. These include Choi, Nam, and Park (1996), quality assurance (for process capability indices); Jones, Wortberg, Kreissig, Hammock, and Rocke (1996), engineering; Bajgier (1992), Seppala, Moskowitz, Plante, and Tang (1995) and Liu and Tang (1996), process control; Chao and Huwang (1987), reliability; Coakley (1996), image processing; Bar-Ness and Punt (1996), communications; and Zoubir and Iskander (1996) and Zoubir and Boashash (1998), signal processing. Ames and Muralidhar (1991) and Biddle, Bruton, and Siegel (1990) provide applications in auditing. Robeson (1995) applies the bootstrap in meteorology, Tambour and Zethraeus (1998) in economics, and Tran (1996) in sports medicine. Roy (1994) and Schafer (1992) provide applications in chemistry, Rothery (1985) and Lanyon (1987) in ornithology. Das Peddada and Chang (1992) give an application in physics. Mooney (1996) covers bootstrap applications in political science. Adams, Gurevitch, and Rosenberg (1997) and Shipley (1996) apply the bootstrap to problems in ecology; Andrieu, Caraux, and Gascuel (1997) in evolution; and Aastveit (1990), Felsenstein (1985), Sanderson (1989, 1995), Sitnikova, Rzhetsky, and Nei (1995), Leal and Ott (1993), Tivang, Nienhuis, and Smith (1994), Schork (1992), Zharkikh and Li (1992, 1995) in genetics. Lunneborg (1987) gives us applications in the behavioral sciences. Abel and Berger (1986) and Brey (1990) give applications in biology. Aegerter, Muller, Nakache, and Boue (1994), Baker and Chu (1990), Barlow and Sun (1989), Mapleson (1986), Tsodikov, Hasenclever, and Loeffler (1998), and Wahrendorf and Brown (1980) apply the bootstrap to a variety of medical problems.

The first monograph on the bootstrap was Efron (1982a). In the 1990s there were a number of books introduced that are dedicated to bootstrapping and/or

related resampling methods. These include Beran and Ducharme (1991), Chernick (1999), Davison and Hinkley (1997), Efron and Tibshirani (1993), Hall (1992a), Helmers (1991b), Hjorth (1994), Janas (1993), Mammen (1992b), Manly (1997), Mooney and Duval (1993), Shao and Tu (1995), and Westfall and Young (1993). Schervish (1995) devotes a section and Sprent (1998) has a whole chapter on the bootstrap. In addition to the bootstrap chapter, the bootstrap is discussed throughout Sprent (1998) because it is one of a few data-driven statistical methods that are the theme of the text. Chernick and Friis (2002) introduce boofstrapping in a biostatistics text for health science students, Hesterberg, Moore, Monaghan, Clipson and Epstein (2003) is a chapter for an introductory statistics text that covers bootstrap and permutation methods and it has been incorporated as Chapter 18 of Moore, McCabe, Duckworth and Sclove (2003) as well as Chapter 14 of the on-line 5th Edition of Moore and McCabe (2005)

Efron has demonstrated the value of the bootstrap in a number of applied and theoretical contexts. In Efron (1988a), he provides three examples of the value of inference through computer-intensive methods. In Efron (1992b) he shows how the bootstrap has impacted theoretical statistics by raising six basic theoretical questions.

Davison and Hinkley (1997) provide a computer diskette with a library of useful SPLUS functions that can be used to implement bootstrapping in a variety of problems. These routines can be used with the commercial Version 3.3 of SPLUS, and they are described in Chapter 11 of the book. Barbe and Bertail (1995) deal with weighted bootstraps.

Two conferences were held in 1990, one in Michigan and the other in Trier, Germany. These conferences specialized in research developments in bootstrap and related techniques. Proceedings from these conferences were published in LePage and Billard (1992) for the Michigan conference, and in Jockel, Rothe, and Sendler (1992) for the conference in Trier.

In 2003, a portion of an issue of the journal *Statistical Science* was devoted to the bootstrap on its Silver Anniversary. It included articles by Efron, Casella, and others. The text by Lahiri (2003a) covers the dependent cases in detail emphasizing block bootstrap methods for time series and spatial data. It also covers model-based methods and provides some coverage of the independent case. Next to this text, Lahiri (2003a) provides the most recent coverage on bootstrap methods. It provides detailed descriptions of the methodology along with rigorous proofs of important theorems. It also uses simulations for comparison of various methods.

1.5. SUMMARY

In this chapter, I have given a basic explanation of Efron's nonparametric bootstrap. I have followed this up with explanations as to why the procedure can be expected to work in a wide variety of applications and also have given a historical perspective to the development of the bootstrap and its early

acceptance or lack thereof. I have also pointed out some of the sections in subsequent chapters and additional references that will provide more details than the brief discussions given in this chapter.

I have tried to make the discussion casual and friendly with each concept described as simply as possible and each definition stated as clearly as I can make them. However, it was necessary for me to mention some advanced concepts including statistical functionals, influence functions, Edgeworth and Cornish–Fisher expansions, and stationary stochastic processes. All these topics are well covered in the statistical literature on the bootstrap.

Since these concepts involve advanced probability and mathematics for a detailed description, I deliberately avoided such mathematical development to try to keep the text at a level for practitioners who do not have a strong mathematical background. Readers with an advanced mathematical background who might be curious about these concepts can refer to the references given throughout the chapter. In addition, Serfling (1980) is a good advanced text that provides much asymptotic statistical theory.

For the practitioner with less mathematical background, these details are not important. It is important to be aware that such theory exists to justify the use of the bootstrap in various contexts, but a deeper understanding is not necessary and for some it is not desirable.

This approach is really no different from the common practice, in elementary statistics texts, to mention the central limit theorem as justification for the use of the normal distribution to approximate the sampling distribution of sums or averages of random variables without providing any proof of the theorem such as Glivenko–Cantelli or Berry–Esseen or of related concepts such as convergence in distribution, triangular arrays, and Lindeberg–Feller conditions.

CHAPTER 2

Estimation

In this chapter, we deal with problems involving point estimates. Section 2.1 covers the estimation of the bias of an estimator by the bootstrap technique. After showing you how to use the bootstrap to estimate bias in general, we will focus on the important application to the estimation of error rates in the classification problem.

This will require that we first provide you with an introduction to the classification problem and the difficulties with the classical estimation procedures when the training set is small. Another application to classification problems, the determination of a subset of features to be included in the classification rule, will be discussed in Section 8.2.

Section 2.2 explains how to bootstrap to obtain point estimates of location and dispersion parameters. When the distributions have finite second moments, the mean and the standard deviation are the common measures. However, we sometimes have to deal with distributions that do not even have first moments (the Cauchy distribution is one such example).

Such distributions come up in practice when taking ratios or reciprocals of random variables where the random variable in the denominator can take on the value zero or values close to zero. The commonly used location parameter is the median, and the interquartile range R is a common measure of dispersion where $R = L_{75} - L_{25}$ for L_{75} the 75th percentile of the distribution and L_{25} the 25th percentile of the distribution.

2.1. ESTIMATING BIAS

2.1.1. How to Do It by Bootstrapping

Let $E(X)$ denote the expected (or mean) value of a random variable X. For an estimator $\hat{\theta}$ of a parameter θ, we consider the random variable $\hat{\theta} - \theta$ for

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

our X. The bias of an estimator $\hat{\theta}$ for θ is defined to be $b = E(\hat{\theta} - \theta)$. As an example, the sample variance,

$$S^2 = \sum_{i-1}^{n} \frac{(X_i - \bar{X})^2}{n-1},$$

based on a sample of n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ from a population distribution with a finite variance, is an unbiased estimator for σ^2, the population variance where

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}.$$

On the other hand, for Gaussian populations the maximum likelihood estimator for σ^2 is equal to

$$(n-1)S^2/n.$$

It is a biased estimator with the bias equal to

$$-\sigma^2/n \qquad \text{since} \qquad E[(n-1)S^2/n] = (n-1)\sigma^2/n.$$

The bootstrap estimator B^* of b is then $E(\theta^* - \hat{\theta})$, where $\hat{\theta}$ is an estimate of θ based on a bootstrap sample. A Monte Carlo approximation to B^* is obtained by doing k bootstrap replications as described in Section 1.1.

For the ith bootstrap replication, we denote the estimate of θ by θ_i^*. The Monte Carlo approximation to B^* is the average of the differences between the bootstrap sample estimates θ_i^* of θ and the original sample estimate $\hat{\theta}$,

$$B_{\text{Monte}} = \sum_{i=1}^{k} (\theta_i^* - \hat{\theta})/k.$$

Generally, the purpose of estimating bias is to improve a biased estimator by subtracting an estimate of its bias from it. In Section 2.1.2, we shall see that Efron's definition of the bias is given by the negative of the definition given here [i.e., $B^* = E(\hat{\theta} - \theta^*)$], and consequently we will add the bias to the estimator rather than subtract it.

Bias correction was the original idea that led to a related resampling method, the jackknife [dating back to Quenouille (1949) and Tukey (1958)]. In the next section, we find an example of an estimator which in small samples has a large bias but not a very large variance. For this problem, the estimation of the prediction error rate in linear discriminant analysis, the bootstrap bias correction approach to the estimating the error rate is a spectacular success!

2.1.2. Error Rate Estimation in Discrimination

First you'll be given a brief description of the two-class discrimination problem. Then, some of the traditional procedures for estimating the expected conditional error rate (i.e., the expected error rate given a training set) will be described. Next we will provide a description some of the various bootstrap-type estimators that have been applied. Finally, results are summarized for some of the simulation studies that compared the bootstrap estimators with the resubstitution and leave-one-out (or cross-validation) estimators.

I again emphasize that this particular example is one of the big success stories for the bootstrap. It is a case where there is strong empirical evidence for the superiority of bootstrap estimates over traditional methods, particularly when the sample sizes are small!

In the two-class discrimination problem you are given two classes of objects. A common example is the case of a target and some decoys that are made to look like the target. The data consist of a set of values for variables which are usually referred to as features.

We hope that the values of the features for the decoys will be different from the values for the targets. We shall also assume that we have a training set (i.e., a sample of features for decoys and a separate sample of features for targets where we know which values correspond to targets and which correspond to decoys). We need the training set in order to learn something about the unknown feature distributions for the target and the decoy.

We shall briefly mention some of the theory for the two-class problem. The interested reader may want to consult Duda and Hart (1973), Srivastava and Carter (1983, pp. 231–253), Fukunaga (1990), or McLachlan (1992) for more details.

Before considering the use of training data, for simplicity, let us suppose that we know exactly the probability density of the feature vector for the decoys and also for the targets. These densities shall be referred to as the class-conditional densities.

Now suppose someone discovers a new object and does not know whether it is a target or a decoy but does have measured or derived values for that object's features. Based on the features, we want to decide whether it is a target or a decoy.

This is a classical multivariate hypothesis testing problem. There are two possible decisions: (1) to classify the object as a decoy and (2) to classify the object as a target. Associated with each possible decision is a possible error: We can decide (1) when the object is a target or we can decide (2) when the object is a decoy.

Generally, there are costs associated with making the wrong decisions. These costs need not be equal. If the costs are equal, Bayes' theorem provides us with the decision rule that minimizes the cost.

For the reader who is not familiar with Bayes' theorem, it will be presented in the context of this problem, after we define all the necessary terms. Even

with unequal costs, we can use Bayes' theorem to construct the decision rule which minimizes the expected cost. This rule is called the Bayes rule and it follows our intuition.

For equal costs, we classify the object as a decoy if the a posteriori probability of a decoy given that we observe the feature vector **x** is higher for the decoy than the a posteriori probability of a target given that we observe feature **x**. We classify it as a target otherwise.

Bayes' theorem gives us a way to compute these a posteriori probabilities. If our a priori probabilities are equal (i.e., before collecting the data we assume that the object is as likely to be a target as it is to be a decoy), the Bayes' rule is equivalent to the likelihood ratio test.

The likelihood ratio test classifies the object as the type which has the greater likelihood for **x** (i.e., the larger class conditional density). For more discussion see Duda and Hart (1973, p. 16).

Many real problems have unequal a priori probabilities; sometimes we can determine these probabilities. In the target versus decoy example, we may have intelligence information that the enemy will put out nine decoys for every real target. In that case, the a priori probability for a target is .1, whereas the a priori probability for a decoy is 0.9.

Let $P_D(\mathbf{x})$ be the class conditional density for decoys and let $P_T(\mathbf{x})$ be the class conditional density for targets. Let C_1 be cost of classifying a decoy as a target, C_2 the cost of classifying a target as a decoy, P_1 the a priori probability for a target and P_2 the a priori probability for a decoy.

Let $P(D|\mathbf{x})$ and $P(T|\mathbf{x})$ denote, respectively, the probability that an object with feature vector **x** is a decoy and the probability that an object with feature vector **x** is a target. For the two-class problem it is obvious that $P(T|\mathbf{x}) = 1 - P(D|\mathbf{x})$ since the object must be one of these two types. By the same argument, $P_1 = 1 - P_2$ for the two-class problem. Bayes' theorem states that

$$P(D|\mathbf{x}) = P_D(\mathbf{x})P_2/[P_D(\mathbf{x})P_2 + P_2(\mathbf{x})P_1] = P_D(\mathbf{x})P_2/[P_D(\mathbf{x})P_2 + P_T(\mathbf{x})(1-P_2)].$$

The Bayes rule, which minimizes expected cost, is defined as follows:

$$\text{Classify the object as a decoy if} \quad \frac{P_D(\mathbf{x})}{P_T(\mathbf{x})} > K,$$

$$\text{Classify the object as a target if} \quad \frac{P_D(\mathbf{x})}{P_T(\mathbf{x})} \leq K,$$

where $K = (C_2P_1)/(C_1P_2)$. See Duda and Hart (1973, pp. 10–15) for a derivation of this result.

Notice that we have made no assumptions about the form of the class-conditional densities. The Bayes rule works for any probability densities. Of course, the form of the decision boundary and the associated error rates

depend on these *known* densities. If we make the further assumption that the densities are both multivariate Gaussian with different covariance matrices, then Bayes' rule has a quadratic decision boundary (i.e., the boundary is a quadratic function of **x**).

If the densities are Gaussian and the covariance matrices are equal, then Bayes' rule has a linear boundary (i.e., the boundary is a linear function of **x**). Both of these results are derived in Duda and Hart (1973, pp. 22–31). The possible decision boundaries for Gaussian distributions with unequal covariances and two-dimensional feature vectors are illustrated in Figure 2.1, which was taken from Duda and Hart (1973, p. 31).

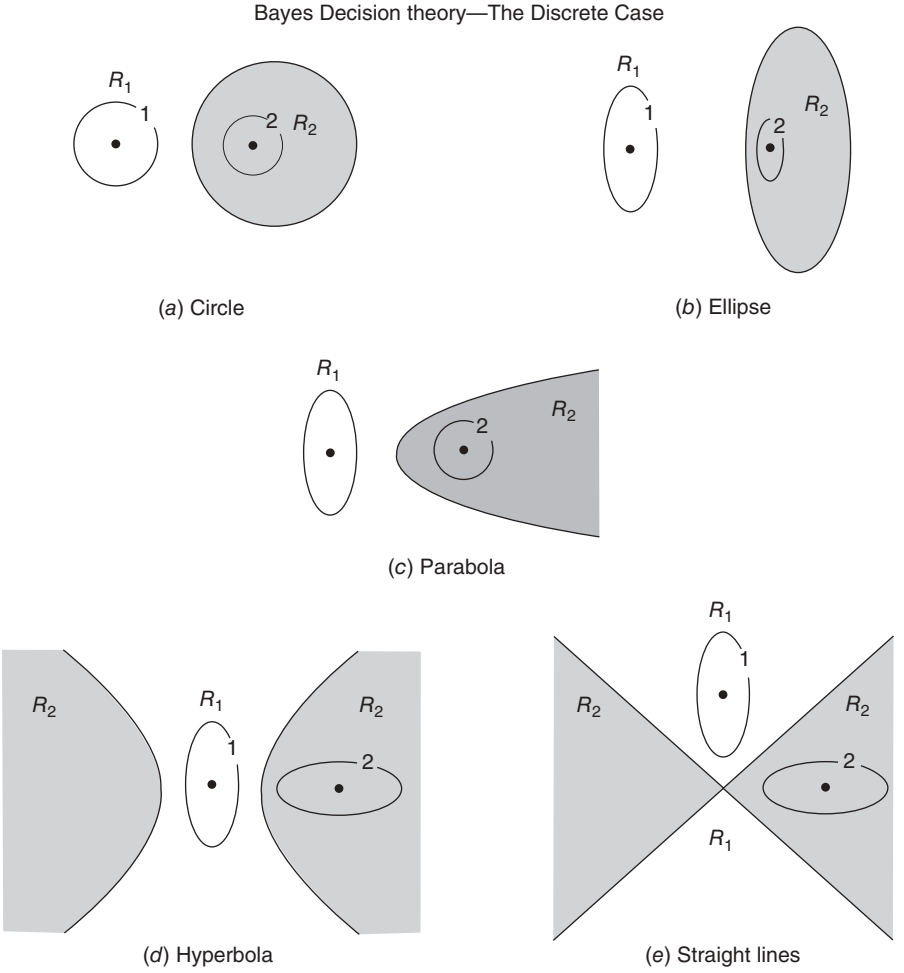


Figure 2.1 Forms for decision boundaries for the general bivariate normal case. [From Duda and Hart (1973), p. 31, with permission from John Wiley and Sons, Inc.]

The circles and ellipses in the figure represent, say, the one sigma equal probability contours corresponding to the covariances. These covariances are taken to be diagonal without any loss of generality. The shaded region R_2 is the region in which class 2 is accepted.

In many practical problems the class-conditional densities are not known. If we assume the densities to be Gaussian, the training samples can be used to estimate the mean vectors and covariance matrices (i.e., the parameters required to determine the densities). If we have no knowledge of the form of the underlying densities, we may use available data whose classes are known (such data are referred to as the training data) to obtain density estimates.

One common approach is to use the kernel density estimation procedure. The rule used in practice replaces the Bayes rule (which is not known) with an approximation to it based on the replacement of the class-conditional densities in the Bayes rule with the estimated densities.

Although the resulting rule does not have the optimal properties of the Bayes rule, we argue that it is an appropriately optimal rule since as the training set gets larger and larger for both classes the estimated densities come closer and closer to the true densities and the rule comes closer and closer to the Bayes rule. For small sample sizes, it at least appears to be a reasonable approach. We shall call this procedure the estimated decision rule. To learn more about kernel discrimination, consult Hand (1981, 1982).

For known class-conditional densities, the Bayes rule can be applied and the error rates calculated by integrating these densities in the region in which a misclassification would occur. In parametric problems, so-called "plug-in" methods compute these integrals using the estimated densities obtained by plugging in the parameter estimates for their unknown values. These plug-in estimates of the error rates are known to be optimistically biased [i.e., they tend to underestimate the actual expected error rates; see Hills (1966)].

When we are unable to make any parametric assumptions, a naive approach is to take the estimated decision rule, apply it to the training data, and then count how many errors of each type would be made. We divide the number of misclassified objects in each class by their respective number of training samples to get our estimates of the error rates. This procedure is referred to as the resubstitution method, since we are substituting training samples for possible future cases and these training samples were already used to construct the decision rule.

In small to moderate sample sizes the resubstitution estimator is generally a poor estimator because it also tends to have a large optimistic bias (actually the magnitude of bias depends on the true error rate). Intuitively, the optimistic bias of the plug-in and resubstitution estimators is due to the fact that in both cases the training data are used to construct the rule and then reused to estimate the error rates.

Ideally, it would be better to estimate the error rates based on an independent set of data with known classes. This, however, creates a dilemma. It is

wasteful to throw away the information in the independent set, since these data could be used to enlarge the training set and hence provide better estimates of the class-conditional densities. On the other hand, the holdout estimator, obtained by the separation of this independent data set for error rate estimation from the training set, eliminates the optimistic bias of resubstitution.

Lachenbruch (1967) [see also Lachenbruch, and Mickey (1968)] provided the leave-one-out estimate to overcome the dilemma. Each training vector is used in the construction of the rule. To estimate the error rate, the rule is reconstructed n times where n is the total number of training vectors. In the ith reconstruction, the ith training vector is left out of the construction.

We then count the ith vector as misclassified if the reconstructed rule would misclassify it. We take the total number misclassified in each class and divide by the number in the respective class to obtain the error rates.

This procedure is referred to as leave-one-out or cross-validation and the estimators are called the leave-one-out estimates or U estimates. Because the observations are left out one at a time, some have referred to it as the jackknife estimator, but Efron (1982a, pp. 53–58) defines another bias correction estimator to be the jackknife estimator [see also Efron (1983)].

Now, you'll be shown how to bootstrap in this application. Essentially, we will apply the bootstrap bias correction procedure that we learned about in Section 2.1.1 to the resubstitution estimator.

The resubstitution estimator, although generally poor in small samples, has a large bias that can be estimated by bootstrapping [see, for example, Chernick, Murthy, and Nealy (1985, 1986)]. Cross-validation (i.e., the leave-one-out estimator) suffers from a large variance for small training sample sizes. Despite this large variance, cross-validation has been traditionally the method of choice.

Glick (1978) was one of the first to recognize the problem of large variance with the leave-one-out estimate, and he proposed certain "smooth" estimators as an alternative. Glick's approach has since been followed up by Snapinn and Knoke (1984, 1985a).

Efron (1982a, 1983) showed that the bootstrap bias correction can produce an estimator that is nearly unbiased (the bias is small though not quite as small as for the leave-one-out estimator) and has a far smaller variance than the leave-one-out estimator. Consequently, the bootstrap is superior in terms of mean square error (a common measure of statistical accuracy).

As a guideline to the practitioner, I believe that the simulation studies to date indicate that for most applications, the .632 estimator is to be preferred. What follows is a description of the research studies to date that provide the evidence to support this general guideline. We shall now describe the various bootstrap estimators that were studied in Efron (1983) and in Chernick, Murthy, and Nealy (1985, 1986, 1988a,b).

It is important to clarify here what error rate we are estimating. It was pointed out by Sorum (1972) that when training data are involved, there are

at least three possible error rates to consider [see also Page (1985) for a more recent account].

In the simulation studies that we review here, only one error rate is considered. It is the expected error rate conditioned on the training set of size n. This averages the two error rates (weighing each equally). It is the natural estimator to consider since in the classification problem the training set is fixed and we need to predict the class for new objects based solely on our prior knowledge and the particular training set at hand.

A slightly different and less appropriate error rate would be the one obtained by averaging these conditional error rates over the distribution of possible training sets of size n. Without carefully defining the error rate to be estimated, confusion can arise and some comparisons may be inappropriate.

The resubstitution estimator and cross-validation have already been defined. The standard bootstrap (obtained using 100–200 bootstrap samples in the simulations of Efron and Chernick, Murthy, and Nealy) uses the bootstrap sample analog to Equation 2.10 of Efron (1983, p. 317) to correct the bias. Define the estimated bias as

$$\omega_h = E_*[\Sigma_i(n^{-1} - P_i^*)Q[y_i, \eta(t_i, X^*)]],$$

where E_* denotes the expectation under the bootstrap random sampling mechanism (i.e., sampling with replacement from the empirical distribution), $Q[y_i, \eta(t_i, X^*)]$ is the indicator function defined to be equal to one if $y_i = \eta(t_i, X^*)$ and zero if $y_i \neq \eta(t_i, X^*)$, y_i is the ith observation of the response, t_i is the vector of predictor variables, and η is the prediction rule.

$\mathbf{X}^*$ is the vector for a bootstrap sample a (of length n) and P_i^* is the ith repetition frequency (i.e. the proportion of cases in particular where the ith sample value occurs). The bootstrap estimate is then $e_{\text{boot}} = \text{err}_{\text{app}} + \omega_h$, where ω_h is the bootstrap estimate of the bias as define above.

This is technically slightly different from the simple bias correction procedure described in Section 2.1.1 but is essentially the same. Using the convention given in Efron (1983), this bias estimate is then added to the apparent error rate to produce the bootstrap estimate.

To be more explicit, let $X_1, X_2, \ldots, X_n$ denote the n training vectors where, say for convenience, $n = 2m$ for m an integer $X_1, X_2, \ldots, X_m$ come from class 1 and $X_{m+1}, X_{m+2}, \ldots, X_n$ come from class 2. A bootstrap sample is generated by sampling with replacement from the empirical distribution for the pooled data $X_1, X_2 \ldots, X_n$.

Although different, this is almost the same as taking m samples with replacement from $X_1, X_2, \ldots, X_m$ and another m samples with replacement from X_{m+1}, $X_{m+2}, \ldots, X_n$. In the latter case, each bootstrap sample contains m vectors from each class, whereas in the former case the number in each class varies according to a binomial distribution where N_1, the number from class 1, is binomial with parameters n and p (with $p = 1/2$) and N_2, the number from class 2, equals $n - N_1$. $E(N_1) = E(N_2) = n/2 = m$.

The approach used in Efron (1983) and Chernick, Murthy, and Nealy (1985, 1986, 1988a,b) is essentially the former approach except that the original training set itself is also selected in the same way as the bootstrap samples. So, for example, when $n = 14$, it is possible to have 7 training vectors from class 1 and 7 from class 2, but also we may have 6 from class 1 and 8 from class 2, and so on.

Once a bootstrap sample has been selected, we treat the bootstrap sample as though it were the training set. We construct the discriminant rule (linear for the simulations under discussion, but the procedure can apply to other forms such as quadratic) based on the bootstrap sample and subtracting the fraction of the observations in the bootstrap sample that would be misclassified by the same rule (where each observation is counted as many times as it occurs in the bootstrap sample).

The first term is a bootstrap sample estimate of the "true" error rate, while the second term is a bootstrap sample estimate of the apparent error rate. The difference is a bootstrap sample estimate of the optimistic bias in the apparent error rate. Averaging these estimates over the k Monte Carlo replications provides a Monte Carlo approximation to the bootstrap estimator.

An explicit formula for the bootstrap estimator and its Monte Carlo approximation is given on p. 317 of Efron (1983). Although the formulas are explicit, the notation is complicated. Nevertheless, the Monte Carlo approximation is simple to describe as we have done above.

The e_0 estimator was introduced as a variant to the bootstrap in Chatterjee and Chatterjee (1983), although the name e_0 came later in Efron (1983). For the e_0 estimate we simply count the total number of training vectors misclassified in each bootstrap sample. The estimate is then obtained by summing over all bootstrap samples and dividing by the total number of training vectors not included in the bootstrap samples.

The .632 estimator is obtained by the formula

$$\mathrm{err}_{632} = 0.368\mathrm{err}_{\mathrm{app}} + 0.632e_0,$$

where $\mathrm{err}_{\mathrm{app}}$ denotes the apparent error rate and e_0 is as defined in the previous paragraph. With only the exception of the very heavy-tailed distributions, the .632 estimator is the clear-cut winner over the other variants.

Some heuristic justification for this is given in Efron (1983) [see also Chernick and Murthy (1985)]. Basically, the .632 estimator appropriately balances the optimistic bias of the apparent error rate with the pessimistic bias of e_0. The reason for this weighting is that 0.368 is a decimal approximation to $1/e$, which is the asymptotic expected percentage of training vectors that are not included in a bootstrap sample.

Chernick, Murthy, and Nealy (1985) devised a variant called the MC estimator. This estimator is obtained just as the standard bootstrap. The difference is that a controlled bootstrap sample is generated in place of the ordinary bootstrap sample. In this procedure, the sample is restricted to include obser-

vations with replication frequencies as close as possible to the asymptotic expected replication frequency.

Another variant, also due to Chernick, Murthy, and Nealy (1985), is the convex bootstrap. In the convex bootstrap, the bootstrap sample contains linear combinations of the observation vectors. This smoothes out the sampling distribution for the bootstrap estimate by allowing a continuum of possible observations instead of just the original discrete set.

A theoretical difficulty with the convex bootstrap is that the bootstrap distribution does not converge to the true distribution since the observations are weighting according to λ which is chosen uniformly on [0,1]. This means that the "resamples" will not behave in large samples exactly like the original samples from the class-conditional densities. We can therefore not expect the estimated error rates to be correct for the given classification rule.

To avoid the inconsistency problem, Chernick, Murthy, and Nealy (1988b) introduced a modified convex bootstrap that concentrates the weight closer and closer to one of the samples, as the training sample size n increases. They also introduced a modification to the .632 estimator which they called the adaptive 632.

It was hoped that the modification of adapting the weights would improve the .632 estimator and increase its applicability, but results were disappointing. Efron and Tibshirani (1997a) introduce .632+, which also modifies the .632 estimator so that it works well for an even wider class of classification problems and a variety of class-conditional densities.

In Efron (1983) other variants—the double bootstrap, the randomized bootstrap, and the randomized double bootstrap—are also considered. The reader is referred to Efron (1983) for the formal definitions of these estimators. Of these, only the randomized bootstrap showed significant improvement over the ordinary bootstrap, and so these other variants were not considered. Follow-up studies did not include the randomized bootstrap.

The randomized bootstrap applies only to the two-class problem. The idea behind the randomized bootstrap is the modification of the empirical distributions for each class by allowing for the possibility that the observed training vectors for class 1 come from class 2 and vice versa.

Efron allowed a probability of occurrence of .1 to the opposite class in the simple version. After modifying the empirical distributions, bootstrap sampling is applied to the modified distributions rather than the empirical distributions and the bias is estimated and then corrected for, just as the standard bootstrap. In a way the randomized bootstrap smoothes the empirical distributions, an idea similar in spirit to the convex bootstrap.

Implementation of the randomized bootstrap by Monte Carlo is straightforward. We sample at random from the pooled training set (i.e., training data from both classes are mixed together) and then choose a uniform random number U. If $U \leq .9$, we assign the observation vector to its correct class. If not, we assign it to the opposite class. To learn more about the randomized bootstrap and other variations, see Efron (1983, p. 320).

For Gaussian populations and small training sample sizes (14–29) the .632 estimator is clearly superior in all the studies in which it was considered, namely Efron (1983), Chernick, Murthy, and Nealy (1985, 1986) and Jain, Dubes, and Chen (1987).

A paper by Efron and Tibshirani (1997), which we have already mentioned, looks at the .632 estimator and a variant called .632+. They treat more general classification problems as compared to just the linear (equal covariances) case that we focus on here.

Chernick, Murthy, and Nealy (1988a,b) consider multivariate (two-dimensional, three-dimensional, and five-dimensional) distributions. Uniform, exponential, and Cauchy distributions are considered. The uniform provides shorter than Gaussian tails to the distribution, and the bivariate exponential provides an example of skewness and the autoregressive family of Cauchy distributions provides for heavier-than-Gaussian tails to the distribution.

They found that for the uniform and exponential distributions the .632 estimator is again superior. As long as the tails are not heavy, the .632 estimator provides an appropriate weighting to balance the opposite biases of e_0 and the apparent error rate.

However, for the Cauchy distribution the e_0 no longer has a pessimistic bias and both the e_0 and the convex bootstrap outperform the .632 estimator. They conjectured that the result would generalize to any distributions with heavy tails. They also believe that skewness and other properties of the distribution which cause it to depart from the Gaussian distribution would have little effect on the relative performance of the estimators.

In Chernick, Murthy, and Nealy (1988b), the Pearson VII family of distributions was simulated for a variety of values of the parameter m. The probability density function is defined as

$$f(x) = \frac{\Gamma(M)|\Sigma|^{1/2}}{\Gamma(m-(p/2))\pi^{p/2}[1+(x-\mu)'(x-\mu)]m},$$

where μ is a location vector, Σ is a scaling matrix, m is a parameter that affects the dependence and controls the tail behavior, p is the dimension, and Γ is the gamma function. The symbol $|\ |$ denotes the determinant of the matrix.

The Pearson VII distributions are all elliptically contoured (i.e., contours of constant probability density are ellipses). An elliptic contoured density is a property the Pearson VII family shares with the Gaussian family of distributions. Only $p = 2$ was considered in Chernick, Murthy, and Nealy (1988b). The parameter m was varied from 1.3 to 3.0. For $p = 2$, second moments exist only for m greater than 2.5 and first moments exist only for m greater than 1.5.

Chernick, Murthy, and Nealy (1988b) found that when $m \leq 1.6$, the pattern observed for the Cauchy distributions in Chernick, Murthy, and Nealy (1988a)

pertained; that is, the e_0 and the convex bootstrap were the best. As m decreases from 2.0 to 1.5, the bias of the e_0 estimator decreases and eventually it changes sign (i.e., goes from a pessimistic to an optimistic bias). For m greater than 2.0, the results are similar to the Gaussian and the light-tailed distributions where the .632 estimator is the clear winner.

Table 2.1 is taken from Chernick, Murthy, and Nealy (1988b). It summarizes for various values of m the relative performance of the estimators. The totals represent the number of cases for which the estimators ranked first, second, and third among the seven considered. The cases vary over the range of the "true" error rates that varied from about .05 to .50.

Table 2.2 is a similar summary taken from Chernick, Murthy, and Nealy (1986) which summarizes the results for the various Gaussian cases considered.

Again, we point out that for most applications the .632 estimator is preferred. It is not yet clear whether or not the smoothed estimators are as good as the best bootstrap estimates.

Snapinn and Knoke (1985b) claim that their estimator is better than the .632 estimator. Their study simulated both Gaussian distributions and a few non-Gaussian distributions.

They also show that the bias correction applied to the smoothed estimators by resampling procedures may be as good as their own smoothed estimators. This has not yet been confirmed in the published literature. Some results comparing the Snapinn and Knoke estimators with the .632 bootstrap and some other estimates in two-class cases are found in Hirst (1996).

For very heavy-tailed distributions, our recommendation would be to use the ordinary bootstrap or the convex bootstrap. But how does the practitioner know that the distributions are heavy-tailed? It may sometimes be possible to make such an assessment from knowledge as to how the data are generated for the practitioner to determine something about the nature of the tails of the distribution. One example would be when the data is ratios where the denominator can be close to zero. But in many practical cases it may not be possible.

To be explicit, consider the case where a feature is the ratio of two random variables and the denominator is known to be approximately Gaussian with zero mean; we will know that the feature has a distribution with tails like the Cauchy. This is because such cases are generalizations of the standard Cauchy distribution. It is a known result that the ratio of two independent Gaussian random variables with zero mean and the same variance has the standard Cauchy distribution. The Cauchy distribution is very heavy-tailed, and even the first moment or mean does not exist.

As the sample size becomes larger, it makes little difference which estimator is used, as the various bootstrap estimates and cross-validation are asymptotically equivalent (with the exception of the convex bootstrap). Even the apparent error rate may work well in very large samples where its bias is much reduced, although never zero. Exactly how large is large is difficult to say

Table 2.1 Summary Comparison of Estimators Using Root Mean Square Error (Number of Simulations on Which Estimator Attained Top Three Ranks)

	.632	MC	e_0	Boot	Conv	U	App	Total
				$M = 1.3$				
First	0	0	2	0	10	0	0	12
Second	3	0	0	9	0	0	0	12
Third	0	9	0	1	2	0	0	12
Total	3	9	2	10	12	0	0	36
				$M = 1.5$				
First	6	1	8	5	12	0	1	33
Second	8	4	0	14	7	0	0	33
Third	3	15	2	4	8	0	1	33
Total	17	20	10	23	27	0	2	99
				$M = 1.6$				
First	1	1	2	1	5	0	2	12
Second	4	3	0	5	0	0	0	12
Third	0	4	0	4	4	0	0	12
Total	5	8	2	10	9	0	2	36
				$M = 1.7$				
First	2	1	2	1	2	1	3	12
Second	3	3	1	4	1	0	0	12
Third	4	2	0	3	2	0	1	12
Total	9	6	3	8	5	0	1	36
				$M = 2.0$				
First	18	1	3	0	1	0	7	30
Second	10	4	4	2	5	2	3	30
Third	1	9	3	8	5	0	3	30
Total	29	14	10	10	11	2	13	90
				$M = 2.5$				
First	21	0	8	1	0	0	3	33
Second	10	3	4	5	4	2	5	33
Third	1	13	1	6	10	0	2	33
Total	32	16	13	12	14	2	10	99
				$M = 3.0$				
First	21	0	6	0	0	0	3	30
Second	9	3	5	3	2	2	6	30
Third	0	8	1	8	11	1	1	30
Total	30	11	12	11	13	3	10	90

Source: Chernick, Murthy, and Nealy (1988b).

Table 2.2 Summary Comparison

Rank	.632	MC	E_0	Boot	Conv	*U*	App	Total
First	72	1	29	6	0	0	1	109
Second	21	13	27	23	11	1	13	109
Third	7	20	8	25	37	7	5	109
Total	100	34	64	54	48	8	19	

Source: Chernick, Murthy, and Nealy (1986).

because the known studies have not yet adequately varied the size of the training sample.

2.1.3. Error Rate Estimation: An Illustrative Problem

In this problem, we have five bivariate normal training vectors from class 1 and have 5 from class 2. For class 1, the mean vector is $\begin{pmatrix}0\\0\end{pmatrix}$ and the covariance matrix is

$$\begin{bmatrix}10\\01\end{bmatrix}.$$

For class 2, the mean vector is $\begin{pmatrix}1\\1\end{pmatrix}$ and the covariance matrix is also

$$\begin{bmatrix}10\\01\end{bmatrix}.$$

The training vectors generated by random sampling from the above distributions are as follows:

For Class 1

$$\begin{pmatrix}2.052\\0.339\end{pmatrix}, \begin{pmatrix}1.083\\-1.320\end{pmatrix}, \begin{pmatrix}0.083\\-1.524\end{pmatrix}, \begin{pmatrix}1.278\\-0.459\end{pmatrix}, \begin{pmatrix}-1.226\\-0.606\end{pmatrix}.$$

For Class 2

$$\begin{pmatrix}1.307\\2.268\end{pmatrix}, \begin{pmatrix}-0.548\\1.741\end{pmatrix}, \begin{pmatrix}2.498\\0.813\end{pmatrix}, \begin{pmatrix}0.832\\1.409\end{pmatrix}, \begin{pmatrix}1.498\\2.063\end{pmatrix}.$$

We generate four bootstrap samples of size 10 and calculate the standard bootstrap estimate of the error rate. We also calculate e_0 and the apparent error rate in order to compute the .632 estimator. We denote by the indices 1, 2, 3, 4, and 5 the respective five bivariate vectors from class 1 and denote by the indices 6, 7, 8, 9, and 10 the respective five bivariate vectors from class 2. A bootstrap sample can be represented by a random set of 10 indices sampled with replacement from the integers 1 to 10. In this instance, our four bootstrap samples are [**9, 3, 10, 8, 1, 9, 3, 5, 2, 6**], [**1, 5, 7, 9, 9, 9, 2, 3, 3, 9**], [**6, 4, 3, 9, 2, 8, 7, 6, 7, 5**], and [**5, 5, 2, 7, 4, 3, 6, 9, 10, 1**].

Bootstrap sample numbers 1 and 2 have five observations from class 1 and five from class 2, bootstrap sample number 3 has four observations from class 1 and six from class 2, and bootstrap sample number 4 has six observations from class 1 and four from class 2. We also observe that in bootstrap sample number 1, indices 3 and 9 repeat once and indices 4 and 7 do not occur. In bootstrap sample number 2, index 9 occurs three times and index 3 twice while indices 4, 6, and 10 do not appear. In bootstrap sample number 3, indices 6 and 7 are repeated once while 1 and 10 do not appear.

Finally, in bootstrap sample number 4, only index 5 is repeated and index 8 is the only one not to appear. These samples are fairly typical of the behavior of bootstrap samples (i.e., sampling with replacement from a given sample), and they indicate how the bootstrap samples can mimic the variability due to sampling (i.e., the sample-to-sample variability).

Table 2.3 shows how the observations in the bootstrap sample were classified by the classification rule obtained using the bootstrap sample. We see that only in bootstrap samples 1 and 2 were any of the bootstrap observations misclassified. So for bootstrap samples 3 and 4 the bootstrap sample estimate of the apparent error rate is zero. In both bootstrap sample 1 and sample 2, only observation number 1 was misclassified and in each sample, observation number 1 appeared one time. So for these two bootstrap samples the estimate of apparent error is 0.1.

Table 2.4 shows the resubstitution counts for the original sample. Since none of the observations were misclassified, the apparent error rate or resubstitution estimate is also zero.

Table 2.3 Truth Table for the Four Bootstrap Samples

	Sample #1 Classified As		Sample #2 Classified As	
True Class	Class 1	Class 2	Class 1	Class 2
Class 1	4	1	4	1
Class 2	0	5	0	5
	Sample #3 Classified As		Sample #4 Classified As	
True Class	Class 1	Class 2	Class 1	Class 2
Class 1	4	0	6	0
Class 2	0	6	0	4

Table 2.4 Resubstitution Truth Table for Original Data

	Sample #1 Classified As	
True Class	Class 1	Class 2
Class 1	5	0
Class 2	0	5

In the first bootstrap sample, observation number 1 was the one misclassified. Observation numbers 4 and 7 did no appear. They both would have been correctly classified since their discriminant function values were 0.030724 for class 1 and –1.101133 for class 2 for observation 4 and –5.765286 for class 1 and 0.842643 for class 2 for observation 7. Observation 4 is correctly classified as coming from class 1 since its class 1 discriminant function value is larger than its class 2 discriminant function value. Similarly observation 7 is correctly classified as coming from class 2.

In the second bootstrap sample, observation number 1 was misclassified, and observation numbers 4, 6, and 10 were missing. Observation 3 was correctly classified as coming from class 1, and observations 60 and 10 were correctly classified as coming from class 2. Table 2.6 provides the coefficients of the linear discriminant functions for each of the four bootstrap samples.

It is an exercise for the reader to calculate the discriminant function values for observation numbers 4, 6, and 10 to see that the correct classifications would be made with bootstrap sample number 2.

In the third bootstrap sample, none of the bootstrap sample observations were misclassified but observation numbers 1 and 10 were missing. Using Table 2.6, we see that for class 1, observation number 1 has a discriminant function value of –3.8587, whereas for class 2 it has a discriminant function value of 2.6268.

Consequently, observation 1 would have been misclassified by the discrimination rule based on bootstrap sample number 3. The reader may easily check this and also may check that observation 10 would be correctly classified as coming from class 2 since its discriminant function value for class 1 is –9.6767 and 13.1749 for class 2.

In the fourth bootstrap sample, none of the bootstrap sample observations are misclassified and only observation number 8 is missing from the bootstrap sample. We see, however, by again computing the discriminant functions, that observation 8 would be misclassified as coming from class 1 since its class 1 discriminant function value is –2.1756 while its class 2 discriminant function value is –2.4171.

Another interesting point to notice from Table 2.5 is the variability of the coefficients of the linear discriminants. This variability in the estimated coefficients is due to the small sample size. Compare these coefficients with the ones given in Table 2.6 for the original data.

Table 2.5 Linear Discriminant Function Coefficients for Bootstrap Samples

True Class	Constant Term	Variable No. 1	Variable No. 2
	Bootstrap Sample No. 1		
Class 1	–1.793	0.685	–2.066
Class 2	–3.781	1.027	2.979
	Bootstrap Sample No. 2		
Class 1	–1.919	0.367	–2.481
Class 2	–3.353	0.584	3.540
	Bootstrap Sample No. 3		
Class 1	–2.343	0.172	–3.430
Class 2	–6.823	1.340	6.549
	Bootstrap Sample No. 4		
Class 1	–1.707	0.656	–2.592
Class 2	–6.130	0.469	6.008

Table 2.6 Linear Discriminant Function Coefficients for the Original Sample

Class Number	Constant Term	Variable No. 1	Variable No. 2
1	–1.493	0.563	–1.726
2	–4.044	0.574	3.653

The bootstrap samples give us an indication of the variability of the rule. This would otherwise be difficult to see. It also indicates that we can expect a large optimistic bias for resubstitution.
We can now compute the bootstrap estimate of bias:

$$\omega_{\text{boot}} = \frac{(0.1-0.1)+(0.1-0.1)+(0.1-0.)+(0.1-0)}{4} = 0.2/4 = 0.05.$$

Since the apparent error rate is zero, the bootstrap estimate of the error rate is also 0.05.

The e_0 estimate is the average of the four estimates obtained by counting in each bootstrap sample the fraction of the observations that do not appear in the bootstrap sample and that would be misclassified. We see from the results above that these estimates are 0.0, 0.0, 0.5, and 1.0 for bootstrap samples 1, 2, 3, and 4, respectively. This yields an estimated value of 0.375.

Another estimate similar to e_0 but distinctly different is obtained by counting all the observations left out of the bootstrap samples that would have

been misclassified by the bootstrap sample rule and dividing by the total number of observations left out of the bootstrap samples. Since only two of the left-out observations were misclassified and only a total of eight observations were left out, this would give us an estimate of 0.250. This amounts to giving more weight to those bootstrap samples with more observations left out.

For the leave-one-out method, observation 1 would be misclassified as coming from class 2 and observation 8 would be misclassified as coming from class 1. This leads to a leave-one-out estimate of 0.200.

Now the .632 estimator is simply $0.368 \times$ (apparent error rate) $+ 0.632 \times (e_0)$. Since the apparent error rate is zero, the .632 estimate is 0.237.

Since the data were taken from independent Gaussian distributions, each with variance one and with the mean equal to zero for population 1 and with mean equal to one for population 2, the expected error rate for the optimal rule based on the distributions being known is easily calculated to approximately 0.240.

The actual error rate for the classifier based on a training set of size 10 can be expected to be even higher. We note that in this example, the apparent error rate and the bootstrap both underestimate the true error rate, whereas the e_0 overestimates it.

The .632 estimator comes surprisingly close to the optimal error rate and gives clearly a better estimate of the conditional error rate (0.295, discussed below) than the others. The number of bootstrap replications is so small in this numerical example that it should not be taken too seriously. It is simply one numerical illustration of the computations involved. Many more simulations are required to draw conclusions, and thus simulation studies such as the ones already discussed are what we should rely on.

The true conditional error rate given the training set can be calculated by integrating the appropriate Gaussian densities over the regions defined by the discriminant rule based on the original 10 sample observations. An approximation based on Monte Carlo generation of new observations from the two classes, classified by the given rule, yields for a sample size of 1000 new observations (500 from each class) an estimate of 0.295 for this true conditional error rate.

Since (for equal error rates) this Monte Carlo estimator is based on a binomial distribution with parameters $n = 1000$ and $p =$ the true conditional error rate, using $p = .3$, we have that the standard error of this estimate is approximately 0.0145 and an approximate 95% confidence interval for p is [0.266, 0.324]. So our estimate of the true conditional error rate is not very accurate.

If we are really interested in comparing these estimators to the true conditional error rate, we probably should have taken 50,000 Monte Carlo replications to better approximate it. By increasing the sample size by a factor of 50, we decrease the standard error by $\sqrt{50}$, which is a factor slightly greater than 7. Hence, the standard error of the estimate would be about 0.002 and the

confidence interval would be $[p_h - 0.004, p_h + 0.004]$, where p_h is the point estimate of the true conditional error rate based on 50,000 Monte Carlo replications. We get 0.004 as the interval half-width since a 95% confidence interval requires a half-width of 1.96 standard errors (close to 2 standard errors).

The width of the interval would then be less than 0.01 and would be useful for comparison. Again, we should caution the reader that even if the true conditional error rate were close to the .632 estimate, we could not draw a strong conclusion from it because we would be looking at only one .632 estimate, one e_0 estimate, one apparent error rate estimate, and so on. It really takes simulation studies to account for the variability of the estimates for us to make valid comparisons.

2.1.4. Efron's Patch Data Example

Sometimes in making comparisons we are interested in computing the ratio of the two quantities. We are given a set of data that enables us to estimate both quantities, and we are interested in estimating the ratio of two quantities, What is the best way to do this? The natural inclination is to take the ratio of the two estimates. Such estimators are called ratio estimators.

However, statisticians know quite well that if both estimates are unbiased, the ratio estimate will be biased (except for special degenerate cases). To see why this is so, suppose that X is unbiased for $E(Y)$ and that Y is unbiased for μ. Since X is unbiased for θ, $E(X) = \theta$ and since Y is unbiased for μ, $E(Y) = \mu$. Then $\theta/\mu = E(X)/E(Y)$, but this is not $E(X/Y)$, which is the quantity that we are interested in. Let us further suppose that X and Y are statistically independent; then we have

$$E(X/Y) = E(X)E(1/Y) = \theta E(1/Y).$$

The reciprocal function $f(z) = 1/z$ is a convex function and therefore by Jensen's inequality (see Ferguson, 1967, pp. 76–78) implies that $f(E(Y)) = f(\mu) = 1/\mu \leq E(f(Y)) = E(1/Y)$. Consequently, $E(X/Y) = \theta\, E(1/Y) \geq \theta/\mu$. The only instance where equality holds is when Y equals a constant. Otherwise $E(X/Y) > \theta/\mu$ and the bias $B = E(X/Y) - \theta/\mu$ is positive. This bias can be large, and it is natural to try to improve the estimate of the ratio by adjusting for the bias. Ratio estimators are also common in survey sampling [see Cochran (1977) for some examples].

In Efron and Tibshirani (1993) an example of ratio estimator is given in Section 10.3 on pages 126–133. This was a small clinical trial used to show the FDA that a product produced at a new plant is equivalent to the product produced at the old plant where the agency had previously approved the product. In this example the product is a patch that infuses a certain natural hormone into the patient's bloodstream. The trial was a crossover trial involving eight subjects. Each subject was given three different patches: one patch that was manufactured at the old plant containing the hormone, one

manufactured at the new plant containing the hormone, and a third patch (placebo) that contained no hormone.

The purpose of the placebo is to establish a baseline level to compare with the hormone. Presumably the subjects were treated in random order with regard to treatment, and between each treatment an appropriate wash-out period is applied to make sure that there is no lingering effect from the previous treatment.

The FDA has a well-defined criterion for establishing bioequivalence in such trials. They require that the new patch produces hormone levels that are within 20% of the amount produced by the old patch to the placebo. Mathematically, we express this as

$$\theta = [E(\text{new}) - E(\text{old})]/[E(\text{old}) - E(\text{placebo})]$$

and require that

$$|\theta| = [|E(\text{new}) - E(\text{old})|]/[|E(\text{old}) - E(\text{placebo})|] \leq 0.20.$$

So, for the FDA, the pharmaceutical company must show equivalence by rejecting the "null" hypothesis of non-equivalence in favor of the alternative of equivalence. So the null hypothesis is $|\theta| \leq 0.20$, versus the alternative that $|\theta| > 0.20$. This is most commonly done by applying Schurmann's two one-sided t tests. In recent years a two-stage group sequential test can be used with the hope of requiring a smaller total sample size than for the fixed sample size test.

For the ith subject we define z_i = (old patch blood level – placebo blood level) and y_i = (new patch blood level – old patch blood level). The natural estimate of θ is the plug-in estimate y_b/z_b, where y_b is the average of the eight y_i and z_b is the average of the eight z_i. As we have already seen, such a ratio estimator will be biased.

Table 2.7 shows the y and z values. Based on these data, we find that the plug-in estimate for θ is –0.0713, which is considerably less than the 0.20 in absolute value. However, the estimate is considerably biased and we might be able to improve our estimate with an adjustment for bias. The bootstrap can be used to estimate this bias as you have seen previously in the error rate estimation problem. The real problem is one of confidence interval estimation or hypothesis testing, and so the methods presented in Chapter 3 might be more appropriate. Nevertheless, we can see if the bootstrap can provide a better point estimate of the ratio. Efron and Tibshirani (1993) generated 400 bootstrap samples and estimated the bias to be 0.0043. They also estimated the standard error of the estimate, and the ratio of the bias estimate divided by the estimated standard error is only 0.041. This is small enough to indicate that the bias adjustment will not be important.

The patch data example is a case of equivalence of a product as it is manufactured in two different plants. It is also common for pharmaceutical

Table 2.7 Patch Data Summary

Subject	Old – Placebo (z)	New – Old (y)
1	8,406	–1,200
2	2,342	2,601
3	8,187	–2,705
4	8,459	1,982
5	4,795	–1,290
6	3,516	351
7	4,796	–638
8	10,238	–2,719
Average	6,342	–452.3

Source: Efron and Tibshirani (1993, p. 373), with permission from CRC Press, LLC.

companies to make minor changes in approved products since the change may improve the marketability of the product. To get the new product approved, the manufacturer must design a small bioequivalence trial much like the one shown in the patch data example. Recently, bootstrap methods have been developed to test for bioequivalence. There are actually three forms of bioequivalence defined. They are individual bioequivalence, average bioequivalence, and population bioequivalence. Depending on the application, one type may be more appropriate to demonstrate than another. We will give the formal definitions of these forms of bioequivalence and show examples of bootstrap methods for demonstrating individual and population bioequivalence in Chapter 8. The approach to individual bioequivalence was so successful that it has become a recommended approach in an FDA guidance document.

It is important to recognize that although the bootstrap adjustment will reduce the bias of the estimator and can do so substantially when the bias is large, it is not clear whether or not it improves the accuracy of the estimate. If we define the accuracy to be the root mean square error (rms), then since the rms error is the square root of the bias squared plus the variance, there is the possibility that although we decrease the bias, we could also be increasing the variance. If the increase in variance is larger than the decrease in the squared bias, the rms will actually increase. This tradeoff between bias and variance is common in a number of statistical problems including kernel smoothing, kernel density estimation, and the error rate estimation problem that we have seen. Efron and Tibshirani (1993, p. 138) caution about the hazards of bias correction methods.

2.2. ESTIMATING LOCATION AND DISPERSION

In this section, we consider point estimates of location parameters. For distributions with finite first and second moments the population mean is a

natural location parameter. The sample mean is the "best" estimate, and bootstrapping adds nothing to the parametric approach. We shall discuss this briefly.

For distributions without first moments, the median is a more natural parameter to estimate the location of the center of the distribution. Again, the bootstrap adds nothing to the point estimation, but we see in Section 2.2.2 that the bootstrap is useful in estimating standard errors and percentiles, which provide measures of the dispersion and measures of the accuracy of the estimates.

2.2.1. Means and Medians

For population distributions with finite first moments, the mean is a natural measure of central tendency. If the first moment does not exist, sample estimates can still be calculated but they tend to be unstable and they lose their meaning (i.e., the sample mean no longer converges to a population mean as the sample size increases).

One common example that illustrates this point is the standard Cauchy distribution. Given a sample size n from a standard Cauchy distribution, the sample mean is also standard Cauchy. So no matter how large we take n to be, we cannot reduce the variability of the sample mean.

Unlike the Gaussian or exponential distributions that have finite first and second moments and have sample means that converge in probability to the population mean, the Cauchy has a sample mean that does not converge in probability.

For distributions like the Cauchy, the sample median does converge to the population median as the sample size tends to infinity. Hence for such cases the sample median is a more useful estimator of the center of the distribution since the population median of the Cauchy and other heavy-tailed symmetric distributions best represents the "center" of the distribution.

If we know nothing about the population distribution at all, we may want to estimate the median since the population median always exists and is consistently estimated by the sample median regardless of whether or not the mean exists.

How does the bootstrap fit in when estimating a location parameter of a population distribution? In the case of Gaussian or the exponential distributions, the sample mean is the maximum likelihood estimate, is consistent for the population mean, and is the minimum variance unbiased estimate. How can the bootstrap top that?

In fact it cannot. In these cases the bootstrap could be used to estimate the mean but we would find that the bootstrap estimate is nothing but the sample mean itself, which is the average of all bootstrap samples, and the Monte Carlo estimate is just an approximation to the sample mean. It would be silly to bootstrap is such a case.

Nevertheless, for the purpose of developing a statistical theory for the bootstrap, the first asymptotic results were derived for the estimate of the mean when the variance is finite (Singh, 1981; Bickel and Freedman, 1981).

Bootstrapping was designed to estimate the accuracy of estimators. This is accomplished by using the bootstrap samples to estimate the standard deviation and possibly the bias of a particular estimator for problems where such estimates are not easily derived from the sample. In general, bootstrapping is not used to produce a better point estimate.

A notable exception was given in Section 2.1, where bias correction to the apparent error rate actually produced a better point estimate of the error rate. This is, however, an exception to the rule.

In the remainder of the book, we will learn about examples for which estimators are given, but we need to estimate their standard errors or construct confidence regions or test hypotheses about the corresponding population parameters.

For the case of distributions with heavy tails, we may be interested in robust estimates of location (the sample median being one such example). The robust estimators are given (e.g., Winsorized mean, trimmed mean, or sample median).

However, the bootstrap becomes useful as an approach to estimating the standard errors and to obtain confidence intervals for the location parameters based on these robust estimators. Some of the excellent texts that deal with robust statistical procedures are Chatterjee and Hadi (1988), Hampel, Ronchetti, Rousseeuw, and Stahel (1986), and Huber (1981).

2.2.2. Standard Errors and Quartiles

The standard deviation of an estimator (also referred to as the standard error for unbiased estimators) is a commonly used estimate of an estimator's variability. This estimate only has meaning if the distribution of the estimator of interest has a finite second moment. In examples for which the estimator's distribution does not have a finite second moment, the interquartile range (the 75th percentile minus the 25th percentile of the estimator's distribution) is often used as a measure of the variability.

Staudte and Sheather (1990, pp. 83–85) provide an exact calculation for the bootstrap estimate of the standard error of the median [originally derived by Maritz and Jarrett (1978)] and compare it to the Monte Carlo approximation for cell lifetime data (obtained as the absolute differences of seven pairs of independent identically distributed exponential random variables).

We shall review Staudte and Sheather's development and present their results here. For the median, they assume for convenience that the sample size n is odd (i.e., $n = 2m + 1$, for m an integer). This makes the exposition easier but is not a requirement.

Maritz and Jarrett (1978) actually provide explicit results for any n. It is just that the median is defined as the average of the two "middle" values when n is even and as the unique "middle" observation $m + 1$ when n is odd.

The functional representing the median is just $T(F) = F^{-1}(1/2)$, where F is the population cumulative distribution and F^{-1} is its inverse function. The sample median is just $X_{(m+1)}$, where $X_{(i)}$ denotes the ith-order statistic (i.e., the ith observation when ordered from smallest to largest).

An explicit expression for the variance of median of the bootstrap distribution can then be derived based on well-known results about order statistics. Let $X^*_{(1)}, \ldots, X^*_{(n)}$ denote the ordered observations from a bootstrap sample taken from $X_1, \ldots, X_n$. Let $x_{(i)}$ denote the ith smallest observation from the original sample. Let $N^*_i = \#\{j: X^*_j = x_{(i)}\}, i = 1, \ldots, n\}$.

Then it can be shown that $\sum_{i=1}^{k} N^*_i$ has the binomial distribution with parameters n and p, where $p = k/n$. Let P^* denote the probability under bootstrap sampling. It follows that

$$P^*\{X^*_{(m+1)} > x_{(k)}\} = P^*\left\{\sum_{i=1}^{k} N^*_i \le n\right\} = \sum_{j-0}^{n} \binom{n}{j}\binom{k}{n}^j \left(\frac{n-k}{n}\right)^{n-j}.$$

Using well-known relationships between binomial sums and the incomplete beta function, Staudte and Sheather (1990) find, letting $w_k = P^*\{X^*_{(m)} = x_{(k)}\}$, that

$$w_k = \frac{n!}{(m!)^2}\int_{(k-1)/n}^{k/n} (1-y)^m y^m dy$$

and then by simple probability calculations the bootstrap variance of $X^*_{(m+1)}$ is

$$\sum_{k-1}^{n} w_k x_{(k)} - \left(\sum_{k-1}^{n} w_k x_{(k)}\right)^2.$$

This result was first obtained by Maritz and Jarrett (1978) and later independently by Efron (1978). Taking the square root of the above expression, we have explicitly obtained, using properties of the bootstrap distribution for the median, the bootstrap estimate of the standard deviation of the sample median without doing any Monte Carlo approximation.

Table 2.8, taken from Staudte and Sheather (1990, p. 85), shows the results required to compute the standard error for the "sister cell" data set. In the table, p_k plays the role of w_k and the above equation using p_k gives.

$\mathrm{SE}_{\mathrm{BOOT}} = 0.173$. However, if we replace p_k with $\hat{p}_k$, we get 0.167 for a Monte Carlo approximation based on 500 bootstrap samples.

Table 2.8 Comparison of Exact and Monte Carlo Bootstrap Distributions for the Ordered Absolute Differences of Sister Cell Lifetimes

K	1	2	3	4	5	6	7
p_k	0.0102	0.0981	0.2386	0.3062	0.2386	0.0981	0.0102
$\hat{p}_k$	0.01	0.128	0.548		0.208	0.098	0.008
$x_{(k)}$	0.3	0.4	0.5	0.5	0.6	0.9	1.7

Source: Staudte and Sheather (1990, p. 85), with permission from John Wiley & Sons, Inc.

For other estimation problems the Monte Carlo approximation to the bootstrap may be required, since we may not be able to provide explicit calculations as we have just done for the median. The Monte Carlo approximation is straightforward. Let $\hat{\theta}$ be the sample estimate of θ and let $\hat{\theta}_i^*$ be the bootstrap estimate of for the ith bootstrap sample. Given k bootstrap samples, the bootstrap estimate of the standard deviation of the estimator $\hat{\theta}$ is, according to Efron (1982a),

$$SD_b\left\{\frac{1}{k-1}\sum_{j=1}^{k}\left[\theta_i^* - \bar{\theta}^*\right]^2\right\}^{1/2},$$

where $\overline{\theta^*}$ is the average of the bootstrap samples. Instead of $\overline{\theta^*}$, one could equally well use $\hat{\theta}$ itself. The choice of $k - 1$ in the denominator was made as the analog to the unbiased estimate of the standard deviation for a sample. There is no compelling argument for using $k - 1$ instead of k in the formula.

For the interquartile range, one straightforward approach is to order the bootstrap sample estimates from smallest to largest. The bootstrap sample observation that equals the 25th percentile (or an appropriate average of the two bootstrap sample estimates closest to the 25th percentile) is subtracted from the bootstrap sample observation that equals the 75th percentile (or an appropriate average of the two bootstrap sample observations closest to the 75th percentile). Once these bootstrap sample estimates are obtained, bootstrap standard error estimates or other measures of spread for the interquartile range can be determined.

Other estimates of percentiles from a bootstrap distribution can be used to obtain bootstrap confidence intervals and test hypotheses as will be discussed in Chapter 3. Such methods could be applied to get approximate confidence intervals for standard errors, interquartile ranges, or any other parameters that can be estimated from a bootstrap sample (e.g., medians, trimmed means, Winsorized means, M-estimates, or other robust location estimates).

2.3. HISTORICAL NOTES

For the error rate estimation problem there is a great deal of literature. For developments up to 1974 see the survey article by Kanal (1974) and see the extensive bibliography by Toussaint (1974). In addition, for multivariate Gaussian features McLachlan has derived the asymptotic bias of the apparent error rate (i.e., the resubstitution estimate) in McLachlan (1976) and it is not zero!

The bias of plug-in rules under parametric assumptions is discussed in Hills (1966). A collection of articles including some bootstrap work can be found in Choi (1986).

There have been a number of simulation studies showing the superiority of versions of the bootstrap over cross-validation when the training sample size is small. Most of the studies have considered linear discriminant functions (although Jain, Dubes, and Chen consider quadratic discriminants). Most consider the two-class problem with two-dimensional feature vectors.

However, Efron (1982a, 1983) and Chernick, Murthy, and Nealy (1985, 1986, and 1988a) considered five-dimensional feature vectors as well. Also, in Chernick, Murthy, and Nealy (1985, 1986, 1988a) some three-class problems were considered. Chernick, Murthy, and Nealy (1988a,b) were the first to simulate the performance of these bootstrap estimators for linear discriminant functions when the populations were not Gaussian. Hirst (1996) proposes a smoothed estimator (a generalization of the Snapinn and Knoke approach) for cases with three or more classes and provides detailed simulation studies showing the superiority of his method. He also compares .632 with the smoothed estimator of Snapinn and Knoke (1985) in two-class problems.

Chatterjee and Chatterjee (1983) considered only the two-class problem, doing only one-dimensional Gaussian simulations with equal variance. They were, however, the first to consider a variant of the bootstrap which Efron later refers to as e_0 in Efron (1983). They also provided an estimated standard error for their bootstrap error rate estimation.

The smoothed estimators have also been compared with cross-validation by Snapinn and Knoke (1984, 1985a). They show that their estimators have smaller mean square error than cross-validation for small training samples sizes, but unfortunately not much has been published comparing the smoothed estimates with the bootstrap estimates. We are aware of one unpublished study, Snapinn and Knoke (1985b), and some results in Hirst (1996).

In the simulation studies of Efron (1983), Chernick, Murthy, and Nealy (1985, 1986), Chatterjee and Chatterjee (1983), and Jain, Dubes, and Chen (1987), only Gaussian populations were considered.

Only Jain, Dubes, and Chen (1987) considered classifiers other than linear discriminants. They looked at quadratic and nearest-neighbor rules. Performance was measured by mean square error of the conditional expected error rate.

Jain, Dubes, and Chen (1987) and Chatterjee and Chatterjee (1983) also considered confidence intervals and the standard error of the estimators, respectively. Chernick, Murthy, and Nealy (1988a,b), Hirst (1996) and Snapinn and Knoke (1985b) considered certain non-Gaussian populations. The most recent results on the .632 estimator and an enhancement of it called .632+ are given in Efron and Tibshirani (1997a).

McLachlan has done a lot of research in discriminant analysis and particularly on error rate estimation. His survey article (McLachlan, 1986) provides a good review of the issues and the literature including bootstrap results up to 1986. Some of the developments discussed in this chapter appear in McLachlan (1992), where he devotes an entire chapter, (Chapter 10) to the estimation of error rates. It includes a section on bootstrap (pp. 346–360).

An early account of discriminant analysis methods is given in Lachenbruch (1975). Multivariate simulation methods such as those used in studies by Chernick, Murthy, and Nealy are covered in Johnson (1987).

The bootstrap distribution for the median is also discussed in Efron (1982a, Chapter 10, pp. 77–78). Mooney and Duval (1993) discuss the problem of estimating the difference between two medians.

Justification (consistency results) for the bootstrap approach to individual bioequivalence came in Shao, Kübler, and Pigeot (2000). The survey article by Pigeot (2001) is an excellent reference for the advantages and disadvantages of the bootstrap and the jackknife in biomedical research, and it includes coverage of the individual bioequivalence application.

CHAPTER 3

Confidence Sets and Hypothesis Testing

Because of the close relationship between tests of hypotheses and confidence intervals, we include both in this chapter. Section 3.1 deals with "nonparametric" bootstrap confidence intervals (i.e., little or no assumptions are made about the form of the distribution being sampled).

There has also been some work on parametric forms of bootstrap confidence intervals and on methods for reducing or eliminating the use of Monte Carlo replications. We shall not discuss these in this text but do include references to the most relevant work in the historical notes (Section 3.5). Also, the parametric bootstrap is discussed briefly in Chapter 6.

Section 3.1.2 considers the simplest technique, the percentile method. This method works well when the statistic used is a pivotal quantity and has a symmetric distribution [see Efron (1981c, and 1982a)].

The percentile method and various other bootstrap confidence interval estimates require a large number of Monte Carlo replications for the intervals to be both accurate (i.e., be as small as possible for the given confidence level) and nearly exact (i.e., if the procedure were repeated many times the percentage of intervals that would actually include the "true" parameter value is approximately the stated confidence levels).

This essentially states for exactness that the actual confidence level of the interval is approximately the stated level. So, for example, if we construct a 95% confidence interval, we would expect that our procedure would produce intervals that contain the true parameter in 95% of the cases. Such is the definition of a confidence interval.

Unfortunately for "nonparametric" intervals, we cannot generally do this. The best we can hope for is to have approximately the stated coverage. Such

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

intervals will be called approximately correct or almost exact. As the sample size increases and the number of bootstrap Monte Carlo replications increases, we can expect the percentile method to be approximately correct and accurate.

Another method that Hall (1992a) refers to as the percentile method is also mentioned in Section 3.1.2. Hall refers to Efron's percentile method as the "other" percentile method.

For pivotal quantities that do not have symmetric distributions, the intervals can be improved by bias adjustment and acceleration constants. This is the approach taken in Efron (1987) and is the topic of Section 3.1.3.

Another approach that also provides better bootstrap confidence intervals is called bootstrap iteration (or double bootstrap). This approach has been studied in detail by Hall and Martin, among others, and is covered in Section 3.1.4. There we provide a review of research results and the developments from Martin (1990a) and Hall (1992a).

In each of the sections, examples are given to instruct the reader in the proper application of the methods, to illustrate their accuracy and correctness. Important asymptotic results will be mentioned, but we shall not delve into the asymptotic theory.

Section 3.1.5 deals with the bootstrap t method for generating bootstrap-type confidence intervals. In some problems, the bootstrap t method may be appropriate and has better accuracy and correctness than the percentile method. It is easier to implement than methods involving Efron's corrections. It is not as computer intensive as the iterated bootstrap. Consequently, it is popular in practice. We applied it in the Passive Plus DX clinical trial at Pacesetter. So Section 3.1.5 is intended to provide the definition of it so that the reader may apply it. The bootstrap t was introduced by Efron, in his monograph (Efron, 1982a).

In Section 3.2, the reader is shown the connection between confidence intervals and hypothesis tests. This close connection enables the reader to see how a confidence interval for a parameter can be reinterpreted in terms of the acceptance or rejection of a hypothesis test with a null hypothesis that the parameter is a specified value.

The confidence level is directly related to the significance level of the test. Knowing this, the reader will be able to test hypotheses by constructing bootstrap confidence intervals for the parameter.

In Section 3.3, we provide examples of hypothesis tests to illustrate the usefulness of the bootstrap approach. In some cases, we can compare the bootstrap tests with other nonparametric tests including the permutation tests from Good (1994) or Manly (1991, 1997).

Section 3.4 provides an historical perspective on the literature for confidence interval estimation and hypothesis testing using the bootstrap approach.

3.1. CONFIDENCE SETS

Before introducing the various bootstrap-type confidence intervals, we will review what a confidence set or region is and then, in Section 3.1.1, present Hartigan's typical value theorem in order to motivate the percentile method of Section 3.1.2. Section 3.1.3 then explains how refinements can be made to handle asymmetric cases where the percentile method does not work well.

Section 3.1.4 presents bootstrap interation. Bootstrap iteration or double bootstrapping is another approach to confidence intervals that overcomes the deficiencies of the percentile method. In Section 3.1.5, we present the bootstrap t method that also overcomes deficiencies of the percentile method but is simpler and more commonly used in practice than the iterated bootstrap and other bootstrap modifications to the percentile method.

What is a confidence set for a parameter vector? Suppose we have a parameter vector ν that belongs to an n-dimensional Euclidean space (denoted by R^n). A confidence set with confidence coefficient $1 - \alpha$ is a set in R^n determined on the basis of a random sample and having the property that if the random sampling were repeated infinitely many times with a new region generated each time, then $100^* (1 - \alpha)\%$ of the time the region will contain ν.

In the simplest case where the parameter is one-dimensional, the confidence region will be an interval or the union of two or more disjoint intervals.

In parametric families of population distributions involving nuisance parameters (parameters required to uniquely specify the distribution but which are not of interest to the investigator) or when very little is specified about the population distribution, it may not be possible to construct confidence sets which have a confidence coefficient that is exactly $1 - \alpha$ for all possible ν and all possible values of the nuisance parameters [see Bahadur and Savage (1956), for example]. We shall see that the bootstrap percentile method will at least provide us with confidence intervals that have confidence coefficient approaching $1 - \alpha$ as the sample size becomes very large.

If we only assume that the population distribution is symmetric, then the typical value theorem of Hartigan (1969) tells us that subsampling methods (e.g., random subsampling) can provide confidence intervals that are exact (i.e., have confidence coefficient $1 - \alpha$ for finite sample sizes). We shall now describe these subsampling methods and present the typical value theorems.

3.1.1. Typical Value Theorems for M-Estimates

We shall consider the case of independent identically distributed observations from a symmetric distribution on the real line. We denote the n random variables by $X_1, X_2, \ldots, X_n$ and their distribution by F_θ. For any set A let $P_\theta(A)$ denote the probability that a random variable X with distribution F_θ has its

value in the set A. As in Efron (1982a, p. 69) we will assume that F_θ has a symmetric density function $f(\cdot)$ so that

$$P_\theta(A) = \int_{-A} f(x-\theta)\,dx,$$

where

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1, \qquad f(x) \geq 0, \qquad \text{and} \qquad f(-x) = f(x).$$

An M-estimate $\hat{\theta}(x_1, x_2, \ldots, x_n)$ for θ, is any solution to the equation

$$\sum_i \Psi(x_i - t) = 0.$$

Here we assume that the observed data $X_i = x_i$ for $i = 1, 2, \ldots, n$ are fixed while t is the variable to solve for.

We note that in general M-estimates need not be unique. The function Ψ is called the kernel, and Ψ is assumed to be antisymmetric and strictly increasing [i.e., $\Psi(-z) = -\Psi(z)$ and $\Psi(z + h) > \Psi(z)$ for all z and for $h > 0$]. Examples of M-estimates are given in Efron (1982a). For an appropriately chosen functions, Ψ many familiar estimates can be shown to be M-estimates including the sample mean and the sample median.

Consider the set of integers $(1, 2, 3, \ldots, n)$. The number of nonempty subsets of this set is $2^n - 1$. Let S be any one of these non-empty subsets. Let $\hat{\theta}_S$ denote an M-estimate based on only those values x_i for i belonging to S.

Under our assumptions about Ψ these M-estimates will be different for differing choices of S. Now let $I_1, I_2, \ldots, I_2^n$ denote the following partition of the real line:

$$\{I_1 = (-\infty, a_1), I_2 = [a_1, a_2), I_3 = [a_2, a_3), \ldots, I_{2^n-1} = [a_{2^n-2}, a_{2^n-1})\},$$

and

$$I_{2^n} = [a_{2^n-1}, +\infty)$$

where a_1 is the smallest $\hat{\theta}_S$, a_2 is the second smallest $\hat{\theta}_S$, and so on. We now are able to state the first typical value theorem.

Theorem 3.1.1.1. The Typical Value Theorem (Hartigan, 1969). The true value of θ has probability $1/2^n$ of being in the interval I_i for $i = 1, 2, \ldots, 2^n$, where I_i is defined as above.

The proof of this theorem is given in Efron (1982a, pp. 70–71). He attributes the method of proof to the paper by Maritz (1979). The theorem came originally from Hartigan (1969), who attributes it to Tukey and Mallows.

We now define a procedure called random subsampling. Let $S_1, S_2, S_3, \ldots, S_{B-1}$ be $B-1$ of the $2^n - 1$ non-empty subsets of $\{1, 2, \ldots, n\}$ selected at random without replacement and let $I_1, I_2, \ldots, I_B$ be the partition of the real line obtained by ordering the corresponding $\hat{\theta}_s$ values, We then have the following typical value theorem, which can be viewed as a corollary to the previous theorem.

Theorem 3.1.1.2. The true value of θ has probability $1/B$ of being in the interval I_i for $i = 1, 2, \ldots, B$ where I_i is defined as above.

For more details and discussion about these results see Efron (1982a). The important point here is that we know the probability that each interval contains θ.

We can then construct an exact $100(j/B)$ percent confidence region for $1 \leq j \leq B - 1$ by simply combining any j of the intervals. The most sensible approach would be to paste together the j intervals in the "middle" if a two-sided interval is desired.

3.1.2. Percentile Method

The percentile method is the most obvious way to construct a confidence interval for a parameter based on bootstrap estimates. Suppose that $\hat{\theta}_i^*$ is the ith bootstrap estimate from the ith bootstrap sample where each bootstrap sample is of size n. By analogy with the case of random subsampling, we would expect that if we ordered the observations from smallest to largest, we would expect an interval that contains 90% of the $\hat{\theta}_i^*$ to be a 90% confidence interval for θ. The most sensible way to choose the interval that excludes the lowest 5% and the highest 5%.

A bootstrap confidence interval generated this way is called a percentile method confidence interval or, more specifically, Efron's percentile method confidence interval. This result (the exact confidence level) would hold if the typical value theorem applied to bootstrap sample estimates just as it did to random subsample estimates. Remember, we also had the symmetry condition and the estimator had to be an M-estimator in Hartigan's theorem.

Unfortunately, even if the distribution is symmetric and the estimator is an M-estimator as is the case for the sample median of, say, a Cauchy distribution, the bootstrap percentile method would not be exact (i.e., the parameter is contained in the generated intervals in exactly the advertised proportion of intervals as the number of generated cases becomes large).

Efron (1982a, pp. 80–81) shows that for the median, the percentile method provides nearly the same confidence interval as the nonparametric interval based on the binomial distribution. So the percentile method works well in some cases even though it is not exact.

Really, the main difference between random subsampling and bootstrapping is that bootstrapping involves sampling with replacement from the origi-

nal sample whereas random subsampling selects without replacement from the set of all possible subsamples. As the sample size becomes large, the difference in the distribution of the bootstrap estimates and the subsample estimates becomes small. Therefore, we expect the bootstrap percentile interval to be almost the same as the random subsample interval. So the percentile intervals inherit the exactness property of the subsample interval asymptotically (i.e., as the sample size becomes infinitely large).

Unfortunately, in the case of small samples (especially for asymmetric distributions) the percentile method does not work well. But fortunately, there are modifications that will get around the difficulties as we shall see in the next section.

In Chapter 3 of Hall (1992a), several bootstrap confidence intervals are defined. In particular, see Section 3.2 of Hall (1992a). In Hall's notation, F_0 denotes the population distribution, F_1 the empirical distribution and F_2 denotes the distribution of the samples drawn at random and with replacement from F_1.

Let φ_0 be the unknown parameter of interest which is expressible as a functional of the distribution F_0. So $\varphi_0 = \varphi(F_0)$. A theoretical α-level percentile confidence interval for φ_0 (by Hall's definition) is the interval $I_1 = (-\infty, \psi + t_0)$, where t_0 is defined so that

$$P(\varphi_0 \leq \psi + t_0) = \alpha.$$

Alternatively, if we define

$$f_1(F_0, F_1) = I\{\varphi(F_0) \leq \varphi(F_1) + t\} - \alpha,$$

then t_0 is a value of t such that $f_t(F_0, F_1) = 0$.

By analogy, a bootstrap one-sided percentile interval for φ_0 would be obtained by solving the equation

$$f_t(F_1, F_2) = 0 \tag{3.1}$$

since in bootstrapping, F_1 replaces F_0 and F_2 replaces F_1. If $\hat{t}_0$ is a solution to Eq. (3.1), the interval $(-\infty, \varphi(F_2) + (\hat{t}_0)$ is a one-sided bootstrap percentile confidence interval for φ. Here $\varphi(F_2)$ is the bootstrap sample estimate for φ. This is a natural way to define a percentile confidence interval according to Hall. It can easily be approximated by Monte Carlo, but differs from Efron's percentile method. Hall refers to Efron's percentile as the "other" percentile method or the "backwards" percentile method.

3.1.3. Bias Correction and the Acceleration Constant

Efron and Tibshirani (1986, pp. 67–70) describe four methods for constructing approximate confidence intervals for a parameter θ. They provide the assump-

tions required for each method to work well. In going from the first method to the fourth, the assumptions become less restrictive while the methods become more complicated but more generally applicable.

The first method is referred to as the standard method. It is obtained by taking the estimator $\hat{\theta}$ of θ and an estimate of its standard deviation $\hat{\sigma}$. The interval $[\hat{\theta} - \hat{\sigma} z_\alpha, \hat{\theta} + \hat{\sigma} z_\alpha]$ is the standard $100(1 - \alpha)\%$ approximate confidence interval for θ. This method works well if $\hat{\theta}$ has an approximate Gaussian distribution with mean θ and standard deviation σ independent of θ.

The second method is the bootstrap percentile method (Efron's definition) described in Section 3.1.2. It works well, when there exists a monotone transformation $\phi = g(\theta)$, such that $\hat{\phi} = g(\hat{\theta})$ is approximately Gaussian with mean ϕ and standard deviation τ independent of ϕ.

The third method is the bias-corrected bootstrap interval, which we discuss in this section. It works well if the transformation $\hat{\phi} = g(\hat{\theta})$ is approximately Gaussian with mean $\phi - z_0\tau$, where z_0 is the bias correction and τ is the standard deviation of $\hat{\phi}$ that does not depend on ϕ.

The fourth method is the BC_a method, which incorporates an acceleration constant a. For it to work well, $\hat{\phi}$ is approximately Gaussian with mean $\phi - z_0\tau_\phi$, where z_0 is the bias correction and τ_ϕ is the standard deviation of $\hat{\phi}$, which does depend on ϕ as follows: $\tau_\phi = 1 + a\phi$, where a is the acceleration constant to be defined later in this section. These results are summarized in Table 6 of Efron and Tibshirani (1986) and are reproduced in Table 3.1.

Efron and Tibshirani (1986) claim that the percentile method automatically incorporates normalizing transformations. To illustrate the difficulties that can be encountered with the percentile method, they consider the case where θ is the bivariate correlation coefficient from a two-dimensional Gaussian distribution and the sample size is 15.

In this case, there is no monotone transformation g that maps $\hat{\theta}$ into $\hat{\phi}$ with $\hat{\phi}$ Gaussian with mean ϕ and constant variance τ^2 independent of ϕ. For a set of data referred to as the "law school data," Efron and Tibshirani (1986) show that the sample bivariate correlation is 0.776.

Assuming we have bivariate Gaussian data with a sample of size 15 and a sample correlation estimate equal to 0.776, we would find that for a bootstrap sample the probability that the correlation coefficient is less than 0.776 based on the bootstrap estimate is only 0.431.

For any monotone transformation, this would also be the probability that the transformed value of the bootstrap sample correlation is less than the transformed value of the original sample correlation [i.e., $g(0.776)$]. However, for the transformed values to be Gaussian or at least a good approximation to the Gaussian distribution and centered about $g(0.776)$, this probability would have to be 0.500 and not 0.431. Note that for symmetric distributions like the Gaussian, the mean is equal to the median. But we do not see that here for the correlation coefficient.

What we see here is that, at least for some values of θ different from zero, no such transformation will work well. Efron and Tibshirani remedy this

Table 3.1 Four Methods of Setting Approximate Confidence Intervals for a Real Valued Parameter θ

Method	Abbreviation	α-Level Endpoint	Correct if
1. Standard	$\theta_S[\alpha]$	$\hat{\theta} + \hat{\sigma} z^{(\alpha)}$	$\hat{\theta} \approx N(\theta, \sigma^2)$ σ is constant
2. Percentile	$\theta_P[\alpha]$	$\hat{G}^{-1}(\alpha)$	There exists a monotone transformation such that $\hat{\phi} = g(\hat{\theta})$, where, $\phi = g(\theta)$, $\hat{\phi} \approx N(\phi, \tau^2)$ and τ is constant.
3. Bias-corrected	$\theta_{BC}[\alpha]$	$\hat{G}^{-1}([\phi[2z_\alpha + z^{(\alpha)}])$	There exists a monotone transformation such that $\hat{\phi} \approx N(\phi - z_0\tau, \tau^2)$ and z_0 and τ are constant.
4. BC_a	$\theta_{BC_a}[\alpha]$	$\hat{G}^{-1}\left(\phi\left[z_0 + \frac{[z_0 + z^{(\alpha)}]}{1 - a[z_0 + z^{(\alpha)}]}\right]\right)$	There exists a monotone transformation such that $\hat{\phi} \approx N(\phi - z_0\tau_\phi, \tau_\phi^2)$, where $\tau\phi = 1 + a\phi$ and z_0 and a are constant.

Note: Each method is correct under more general assumptions than its predecessor. Methods 2, 3, and 4 are defined in terms of the percentile of G, the bootstrap distribution.

Source: Efron and Tibshirani (1986, Table 6) with permission from The Institute of Mathematical Statistics.

problem by making a bias correction to the percentile method. Basically, the percentile method works if exactly 50% of the bootstrap distribution for $\hat{\theta}$ is less than $\hat{\theta}$.

By applying the Monte Carlo approximation, we determine an approximation to the bootstrap distribution. We find the 50th percentile of this distribution and call it $\hat{\theta}^*_{50}$. Taking this bias B to be $\hat{\theta} - \hat{\theta}^*_{50}$, we see that $\hat{\theta} - B$ equals $\hat{\theta}^*_{50}$ and so B is called the bias correction.

Another way to look at it, which is explicit but may be somewhat confusing, is to define $z_0 = \Phi^{-1}\{\hat{G}(\hat{\theta})\}$ (where Φ^{-1} is the inverse of the cumulative Gaussian distribution and $\hat{G}$ is the cumulative bootstrap sample distribution for θ. For a central $100(1 - 2\alpha)\%$ confidence interval, we then take the lower endpoint to be $\hat{G}^{-1}(\Phi\{2z_0 + z^{(\alpha)}\})$ and the upper endpoint to be $\hat{G}^{-1}(\Phi\{2z_0 + z^{(1-\alpha)}\})$. This is how Efron defines the bias correction method in Efron (1982a) and Efron

and Tibshirani (1986), where $z^{(\alpha)}$ satisfies $\Phi(z^{(\alpha)}) = \alpha$. Note that we use the "hat" notation over the cumulative bootstrap distribution G to indicate that Monte Carlo estimate of it is used.

It turns out that in the case of the law school data (assuming that it is a sample from a bivariate Gaussian distribution) the exact central 90% confidence interval is [0.496, 0.898]. The percentile method gives an interval of [0.536, 0.911] and the bias-corrected method yields [0.488, 0.900]. Since the bias-corrected method comes closer to the exact interval, we can conclude, in this case, that it is better than percentile method for the correlation coefficient.

What is important here is that this bias-correction method will work no matter what the value of θ really is. This means that after the adjustment, the monotone transformation leads to a distribution that is approximately Gaussian and whose variance does not depend on the transformed value, ϕ. If the variance cannot be made independent of ϕ, then a further adjustment, referred to as the acceleration constant a, is required.

Schenker (1985) provides an example for which the bias-correct percentile method did not work very well. It involves a χ^2 random variable with 19 degrees of freedom. In Efron and Tibshirani (1986) and Efron (1987) it is shown that the use of an acceleration constant overcomes the difficulty. It turns out in examples like Schenker's that there is a monotone transformation that works after a bias correction. The problem is that the resulting Gaussian distribution has a standard deviation τ_ϕ that depends linearly on ϕ (i.e., $\tau_\phi = 1 + a\phi$, where a is called the acceleration constant). A difficulty in the application of this modification to the bootstrap is the determination of the acceleration constant, a.

Efron found that a good approximation to the constant is one-sixth of the skewness of the score statistic evaluated at $\hat{\theta}$. See Efron and Tibshirani (1986) for details and examples of the computations involved.

Although this method seems to work in very general cases, it is complicated and may not be necessary. Bootstrap iteration to be explained in Section 3.1.4 is an alternative, as is the bootstrap percentile t method of Section 3.1.5.

These methods have a drawback that they share with the bootstrap percentile t intervals, namely, that they are not monotone in the assumed level of coverage (i.e., one could decrease the confidence level and not necessarily get a shorter interval that is contained in the interval obtained at the higher confidence level). This is not a desirable property and goes counter to our intuition about how confidence intervals should behave.

3.1.4. Iterated Bootstrap

A number of authors have contributed to the literature on bootstrap iteration, and we mention many of these contributors in the historical notes (Section 3.4). Major contributions were made by Peter Hall and his graduate student

Michael Martin. Martin (1990a) provides a clear and up-to-date summary of these advances [see also Hall (1992a, Chapter 3)].

Under certain regularity conditions on the population distributions, there has developed an asymptotic theory for the degree of closeness of the bootstrap confidence intervals to their stated coverage probability. Details can be found in a number of papers [e.g., Hall (1988b), Martin (1990a)].

An approximate confidence interval is said to be first-order accurate if its coverage probability differs from its advertised coverage probability by terms which go to zero at a rate of $n^{-1/2}$. The standard intervals discussed in Section 3.1.3 are first-order accurate. The BC_a intervals of Section 3.1.3 and the iterated bootstrap intervals to be discussed in this section are both second-order accurate (i.e., the difference goes to zero at rate n^{-1}).

A more important property for a confidence interval than just being accurate would be for the interval to be as small as possible for the given coverage probability. It may be possible to construct a confidence interval using one method which has coverage probability of 0.95, and yet it may be possible to find another method to use which will also provide a confidence interval with coverage probability 0.95, but the latter interval is actually shorter!

Confidence intervals that are "optimal" in the sense of being the shortest possible for the given coverage are said to be "correct." Efron (1990) provides a very good discussion of this issue along with some examples.

A nice property of these bootstrap intervals (i.e., the BC_a and the iterated bootstrap) is that in addition to being second-order accurate, they are also close to the ideal of "correct" interval in a number of problems where it makes sense to talk about "correct" intervals.

In fact the theory has gone further to show for certain broad parametric families of distributions that corrections can be made to get third-order accurate (i.e., with rate $n^{-3/2}$) intervals (Hall, 1988; Cox and Reid, 1987a and Welch and Peers, 1963).

Bootstrap iteration provides another way to improve the accuracy of bootstrap confidence intervals. Martin (1990a) discusses the approach of Beran (1987) and shows for one-sided confidence intervals that each bootstrap iteration improves the coverage by a factor of $n^{-1/2}$ and for two-sided intervals by n^{-1}.

What is a bootstrap iteration? Let us now describe the process. Suppose we have a random sample $\mathbf{X}$ of size n with observations denoted by $X_1, X_2, X_3, \ldots, X_n$. Let $X_1^*, X_2^*, X_3^*, \ldots, X_n^*$ denote a bootstrap sample obtained from this sample and let $\mathbf{X}^*$ denote this sample. Let I_0 denote a nominal $1 - \alpha$ level confidence interval for a parameter ϕ of the population from which the original sample was taken. For example, I_0 could be a $1 - \alpha$ level confidence interval for ϕ obtained by Efron's percentile method. To illustrate the dependence of I_0 on the original sample $\mathbf{X}$ and the level $1 - \alpha$, we denote it as $I_0(\alpha|\mathbf{X})$. We then denote the actual coverage of the interval $I_0(\alpha|\mathbf{X})$ by $\pi_0(\alpha)$.

Let β_α be the solution to

$$\pi_0(\beta_\alpha) = P\{\theta \in I_0(\beta_\alpha | \mathbf{X}\} = 1 - \alpha. \tag{3.2}$$

Now let $I_0(\beta_\alpha|\mathbf{X}^*)$ denote the version of I_0 computed using the resample in place of the original sample. The resampling principle of Hall and Martin (1988a) states that to obtain better coverage accuracy than given by the original interval I_0 we use $I_0(\beta_\alpha|\mathbf{X}^*)$ where

$$\hat{\beta}_\alpha \text{ is the estimate of } \beta_\alpha$$

in Equation (3.2) obtained by replacing ϕ with $\hat{\theta}$ and $\mathbf{X}$ with $\mathbf{X}^*$. To iterate again we just use the newly obtained interval in place of I_0 and apply the same procedure to it. An estimate based on a single iteration is called the double bootstrap and is the most common iterated estimate used in practice.

The algorithm just described is theoretically possible but in practice a Monte Carlo approximation must be used. In the Monte Carlo approximation B bootstrap resamples are generated. Details of the bootstrap iterated confidence interval are given in Martin (1990a, pp. 1113–1114). Although it is a complicated procedure to describe the basic idea is that by resampling from the B bootstrap resamples, we can estimate the point β_α and use that estimate to correct the percentile intervals. Results for particular examples using simulations are also given in Martin (1990a).

Clearly, the price paid for this added accuracy in the coverage of the confidence interval is an increase in the number of Monte Carlo replications. If we have an original sample size n and each bootstrap resample is of size n, then the number of replications will be nB_1B_2 where B_1 is the number of bootstrap samples taken from the original sample and B_2 is the number of bootstrap samples taken from each resample. In his example of two-sided intervals for the studentized mean from a folded normal distribution, Martin (1990a) uses $n = 10$, $B_1 = B_2 = 299$. The examples do seem to be in agreement with the asymptotic theory in that a single bootstrap iteration does improve the coverage in all cases considered.

Bootstrap iteration can be applied to any bootstrap confidence interval to improve the rate of convergence to the level $1 - \alpha$. Hall (1992a) remarks that although his version of the percentile method may be more accurate than Efron's, bootstrap iteration works better on Efron's percentile method. The reason is not clear and the observation is based on empirical findings. A single bootstrap iteration provides the same type correction as BC_a does to Efron's percentile method. Using more than one bootstrap iteration is not common practice. This is due to the large increase in complexity and computation compared to the small potential gain in accuracy of the confidence interval.

3.1.5. Bootstrap Percentile *t* Confidence Intervals

The iterated bootstrap method and the BC_a confidence interval both provide improvements over Efron's percentile method, but both are complicated and the iterated bootstrap is even more computer-intensive than other bootstraps. The idea of the bootstrap percentile *t* method is found in Efron (1982a). A clearer presentation can be found in Efron and Tibshirani (1993, pp. 160–167). As a consequence of these attributes, it is popular in practice.

It is a simple method and has higher-order accuracy compared to Efron's percentile method. To be precise, bootstrap percentile *t* confidence intervals are second-order accurate (when they are appropriate). See Efron and Tibshirani (1993, pp. 322–325). Consequently, it is popular in practice. We used it in the Passive Plus DX clinical trial.

We shall now describe it briefly. Suppose that we have a parameter θ and an estimate θ_h for θ. Let θ^* be a nonparametric bootstrap estimate for θ based on a bootstrap sample and let S^* be an estimate of the standard deviation for θ_h based on the bootstrap samples. Define $T^* = (\theta^* - \theta_h)/S^*$. For each of the B bootstrap estimates θ^*, there is a corresponding T^*. We find the percentiles of T^*. For an approximate two-sided $100(1 - 2\alpha)\%$ confidence interval for θ, we take the interval $[\theta_h - t^*_{(1-\alpha)}S, \theta_h - t^*_{(\alpha)}]$, where $t^*_{(1-\alpha)}$ is the $100(1 - \alpha)$ percentile of the T^* values and $t^*_{(\alpha)}$ is the 100α percentile of the T^* values and S is the estimated standard deviation for θ_h. This we call the bootstrap *t* (or bootstrap percentile *t* as Hall refers to it) two-sided $100(1 - 2\alpha)\%$ confidence interval for .

A difficulty with the bootstrap *t* is the need for an estimate of the standard deviation S for θ_h and the corresponding bootstrap estimate S^*. In some problems there are obvious estimates, as in the simple case of a sample mean or the difference between the experimental group and control group means). For more complex parameters (e.g., C_{pk}) S may not be available.

3.2. RELATIONSHIP BETWEEN CONFIDENCE INTERVALS AND TESTS OF HYPOTHESES

In Section 3.1 of Good (1994), hypothesis testing for a single location parameter, θ, of a univariate distribution is introduced. In this it is shown how confidence intervals can be generated based on the hypothesis test. Namely for a $100(1 - \alpha)\%$ confidence interval, you include the values of θ at which you would not reject the null hypotheses at the level α. Conversely, if we have a $100(1 - \alpha)\%$ confidence interval for θ, we can construct an α level hypothesis test by simply accepting the hypothesis that $\theta = \theta_0$ if θ_0 is contained in the $100(1 - \alpha)\%$ confidence interval for θ and rejecting if it is outside of the interval.

In problems involving nuisance parameters, this procedure becomes more complicated. Consider the case of estimating the mean μ of a normal distribution when the variance σ^2 is unknown. The statistic $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ has Student's t distribution with $n-1$ degrees of freedom where

$$\bar{x}=\sum_{i-1}^{n} x_i/n,\ s=\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)}.$$

Here n is the sample size and x_i is the ith observed value. What is nice about the t statistic is that its distribution is independent of the nuisance parameter σ^2 and it is a pivotal quantity. Because its distribution does not depend on σ^2 or any other unknown quantities, we can use the tables of the t distribution to determine probabilities such as $P[a \le t \le b]$, where $t=\frac{\bar{x}-\mu}{s/\sqrt{n}}$.

Now t is also a pivotal quantity, which means that probability statements like the one above can be converted into confidence statements involving the unknown mean, μ. So if

$$P[a \le t \le b] = 1-\alpha, \tag{3.3}$$

then the probability is also $1-\alpha$ that the random interval

$$\left[\bar{x}-\frac{bs}{\sqrt{n}}, \bar{x}-\frac{as}{\sqrt{n}}\right] \tag{3.4}$$

includes the true value of the parameter μ, This random interval is then a $100(1-\alpha)\%$ confidence interval for μ.

The interval (3.4) is a $100(1-\alpha)\%$ confidence interval for μ, and we can start with Eq. (3.3) and get Eq. (3.4) or vice versa. If we are testing the hypothesis that $\mu = \mu_0$ versus the alternative that μ differs from μ_0, using (3.2), we replace μ with μ_0 in the t statistic and reject the hypothesis at the α level of significance if $t < a$ or if $t > b$.

We have seen earlier in this chapter how to construct various bootstrap confidence intervals with confidence level approximately $100(1-\alpha)\%$. Using these bootstrap confidence intervals, we will be able to construct hypothesis tests by rejecting parameter values if and only if they fall outside the confidence interval. In the case of a translation family of distributions, the power of the test for the translation parameter is connected to the width of the confidence interval.

In the next section we shall illustrate the procedure by using a bootstrap confidence interval for the ratio of two variances in order to test the equality of the variances. This one example should suffice to illustrate how bootstrap tests can be obtained.

3.3. HYPOTHESIS TESTING PROBLEMS

In principle, we can use any bootstrap confidence interval for a parameter to construct a hypothesis test just as we have described it in the previous section (as long as we have a pivotal or asymptotically pivotal quantity or have no nuisance parameters). Bootstrap iteration and the use of bias correction with the acceleration constant are two ways by which we can provide more accuracy to the confidence interval by making the interval shorter without increasing the significance level. Consequently, the corresponding hypothesis test based on the iterated bootstrap or BC_a confidence interval will be more powerful than the test based on Efron's percentile interval, and it will more closely maintain the advertised level of the test.

Another key point that relates to accuracy is the choice of a test statistic that is asymptotically pivotal. Fisher and Hall (1990) pointed out that tests based on pivotal statistics often result in significance levels that differ from the advertised level by $O(n^{-2})$ as compared to $O(n^{-1})$ for tests based on non-pivotal statistics.

As an example, Fisher and Hall (1990) show that for the one-way analysis of variance, the F ratio is appropriate for testing equality of means when the variances are equal from group to group. For equal (homogeneous) variances the F ratio test is asymptotically pivotal.

However, when the variances differ (i.e., are heterogeneous) the F ratio depends on these variances, which are nuisance parameters. For the heterogeneous case the F ratio is not asymptotically pivotal. Fisher and Hall use a statistic first proposed by James (1951) which is asymptotically pivotal. Additional work on this topic can be found in James (1954).

In our example, we will be using an F ratio to test for equality of two variances. Under the null hypothesis that the two variances are equal, the F ratio will not depend on the common variance and is therefore pivotal.

In Section 3.3.2 of Good (1994), he points out that permutation tests had not been devised for this problem. On the other hand, there is no problem with bootstrapping. If we have n_1 samples from one population and n_2 from the second, we can independently resample with sample sizes of n_1 and n_2 from population one and population two, respectively.

We construct a bootstrap value for the F ratio by using a bootstrap sample of size n_1 from the sample from population one to calculate the numerator (a sample variance estimate for population one) and a bootstrap sample of size n_2 from the sample from population two to calculate the denominator (a sample variance estimate for population two). Since the two variances are equal under the null hypothesis, we expect the ratio to be close to one. By repeating this many times, we are able to get a Monte Carlo approximation to the bootstrap distribution for the F ratio. This distribution should be centered about one when the null hypothesis is true, and the extremes of the bootstrap distribution tell us how far from one we need to set our threshold

for the test. Since the F ratio is pivotal under the null hypothesis, we use the percentiles of the Monte Carlo approximation to the bootstrap distribution to get critical points from the hypothesis test. Alternatively, we could use the more sophisticated bootstrap confidence intervals, but in this case it is not crucial.

In the above example under the null hypothesis we assume $\sigma_1^2/\sigma_2^2 = 1$, and we would normally reject the null hypothesis in favor of the alternative that $\sigma_1^2/\sigma_2^2 \neq 1$, if the F ratio differs significantly from 1. However, in Hall (1992a, Section 3.12) he points out that the F ratio for the bootstrap sample should be compared or "centered" at the sample estimate rather than at the hypothesized value. Such an approach is known to generally lead to more powerful tests than the approach based on sampling at the hypothesized value. See Hall (1992a) or Hall and Wilson (1991) for more examples and a more detailed discussion of this point.

3.3.1. Tendril DX Lead Clinical Trial Analysis

In 1995 Pacesetter Inc., a St. Jude Medical Company that produces pacemakers and leads for patients with bradycardia, submitted a protocol to the United States Food and Drug Administration (FDA) for a clinical trial to demonstrate the safety and effectiveness of an active fixation steroid eluting lead. The study called for the comparison of the Tendril DX model 1388T with a concurrent control, the market-released Tendril model 1188T active fixation lead.

The two leads are almost identical, with the only differences being the use of titanium nitride on the tip of the 1388T lead and the steroid eluting plug also in the 1388T lead. Both leads were designed for implantation in either the atrial or the ventricular chambers of the heart, to be implanted with dual chamber pacemakers (most commonly Pacesetter's Trilogy DR+ pulse generator).

From the successful clinical trials of a competitor's steroid eluting leads and other research literature, it is known that the steroid drug reduces inflammation at the area of implantation. This inflammation results in an increase in the capture threshold for the pulse generator in the acute phase (usually considered to be the first six months post-implant).

Pacesetter statisticians (myself included) proposed as its primary endpoint for effectiveness a 0.5-volt or greater reduction in the mean capture threshold at the three-month follow-up for patients with 1388T leads implanted in the atrial chamber when they are compared to similar patients with 1188T leads implanted in the atrial chamber. The same hypothesis test was used for the ventricular chamber.

Patients entering the study were randomized as to whether they received the 1388T steroid lead or the 1188T lead. Since the effectiveness of steroid is well established from other studies in the literature, Pacesetter argued that it

would be unfair to patients in the study to give them only a 50–50 chance of receiving the 1388T lead (which is expected to provide less inflammation and discomfort and lower capture thresholds).

So Pacesetter designed the trial to have reasonable power to detect a 0.5-volt improvement and yet give the patient a 3-to-1 chance of receiving the 1388T lead. Such an unbalanced design required more patients for statistical conformation of the hypothesis (i.e., based on Gaussian assumptions, a balanced design required 50 patients in each group, whereas with the 3-to-1 randomization 99 patients were required in the experimental group and 33 in the control group to achieve the same power for the test at the 0.05 significance level), a total of 132 patients compared to the 100 for the balanced design.

The protocol was approved by the FDA and the trial proceeded. Interim reports and a pre-market approval report (PMA) were submitted to the FDA and the leads were approved for market release in June 1997.

Capture thresholds take on very discrete values due to the discrete programmed settings. Since the early data at three months was expected to be convincing but the sample size possibly relatively small, nonparametric approaches were taken as alternatives to the standard t tests based on Gaussian assumptions.

The parametric methods would only be approximately valid for large sample sizes due to the non-Gaussian nature of capture threshold distributions (possibly skewed, discrete and truncated). The Wilcoxon rank sum test was used as the nonparametric standard for showing improvement in the mean (or median) of the capture threshold distribution, and the bootstrap percentile method was also used to test the hypothesis.

Figures 3.1 and 3.3 show the distributions (i.e., histograms) of bipolar capture thresholds for 1188T and 1388T leads in the atrium and the ventricle, respectively, at the three-month follow-up visit. The variable, named "leadloc," refers to the chamber of the heart where the lead was implanted.

Figures 3.2 and 3.4 provide the bootstrap histogram of the difference in mean atrial capture threshold and mean ventricular capture threshold, respectively, for the 1388T leads versus the 1188T leads at the three-month follow-up.

The summary statistics in the box are N, the number of bootstrap replications; Mean, the mean of the sampling distribution; Std Deviation, the standard deviation of the bootstrap samples; Minimum, the smallest values out of the 5000 bootstrap estimates of the mean difference; and Maximum, the largest value out of the 5000 bootstrap estimates of the mean difference. Listed on the figures is the respective number of samples for the control (1188T) leads and for the investigational (1388T) leads in the original sample for which the comparison is made.

It also shows the mean difference of the original data that should be (and is) close in value to the bootstrap estimate of the sample mean. The estimate of the standard deviation for the mean difference is also given on the figures.

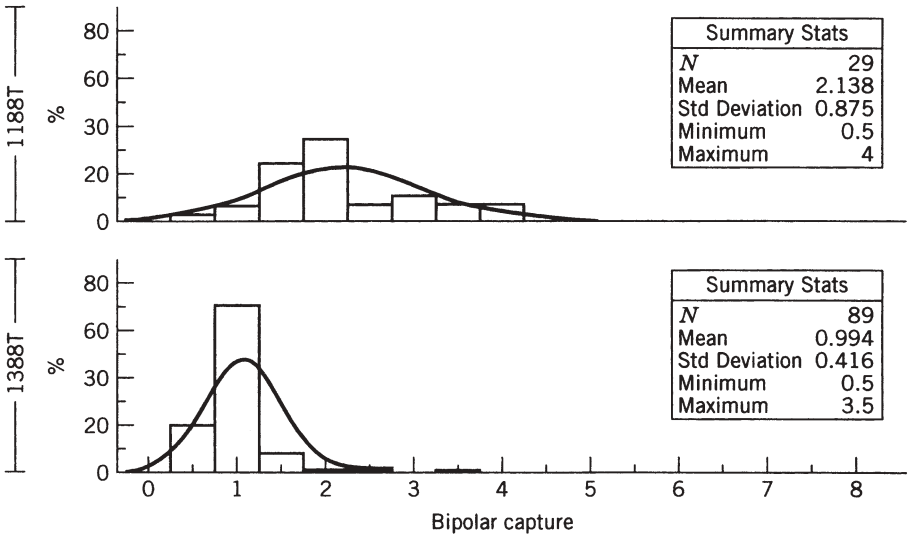


Figure 3.1 Capture threshold distributions for the three-month visit (leadloc; atrial chamber).

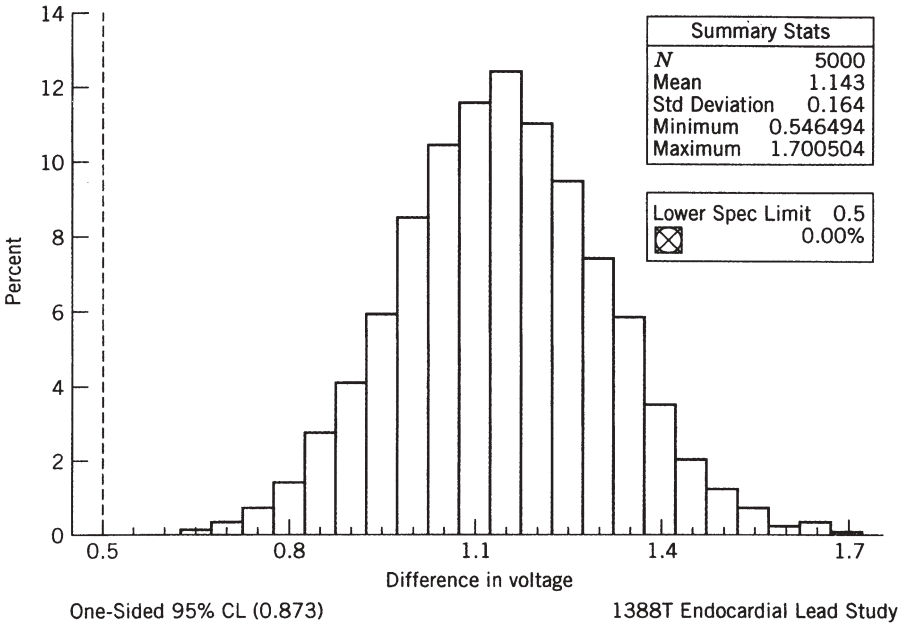


Figure 3.2 Distribution of bootstrapped data sets (atrium) bipolar three-month visit data as of March 15, 1996.

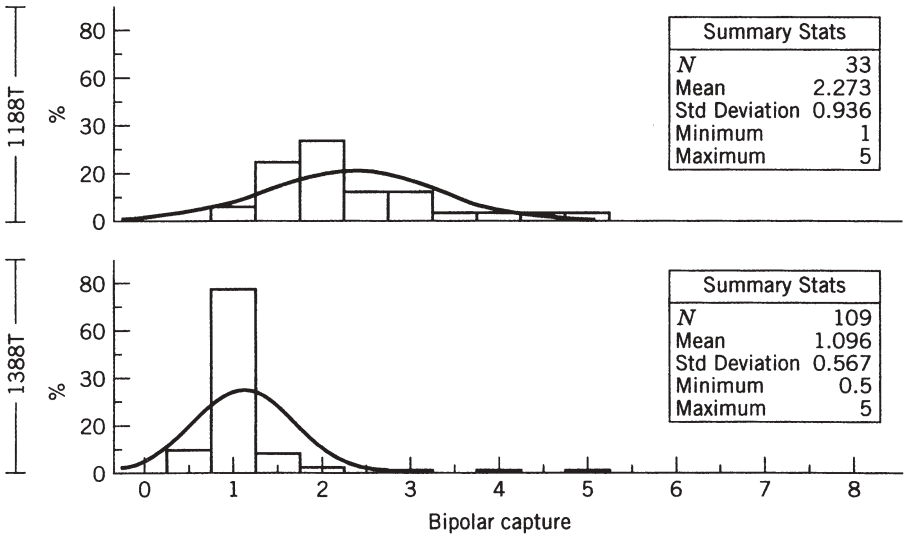


Figure 3.3 Capture threshold distributions for three-month visit (leadloc; ventricular chamber).

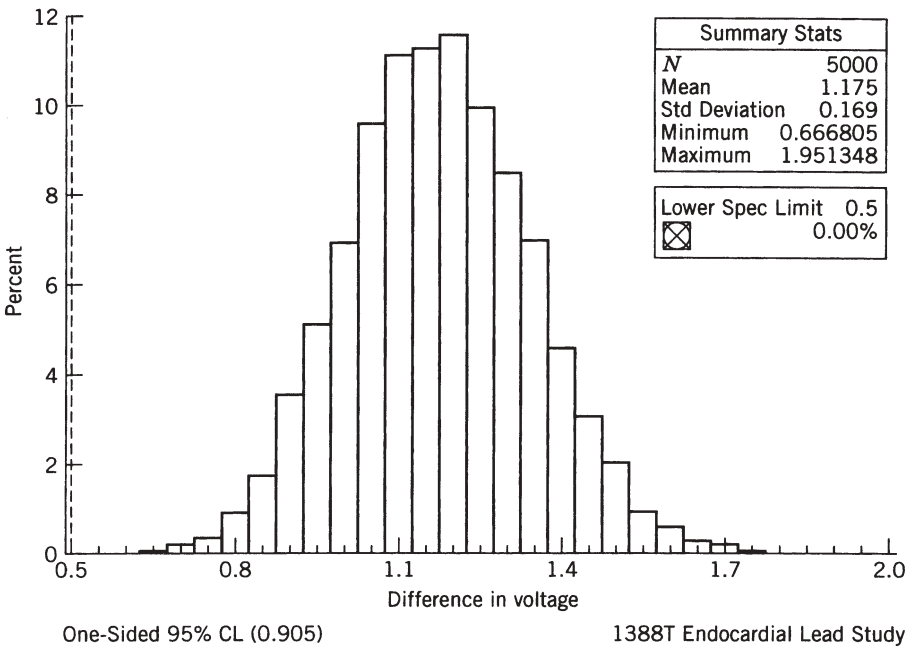


Figure 3.4 Distribution of bootstrapped data sets (ventricle) bipolar three-month data as of March 15, 1996.

We note that this too is very close in value to the bootstrap estimate for these data.

The histograms are based 5000 bootstrap replications on the mean differences. Also shown on the graph of the histogram is the lower 5th percentile (used in Efron's percentile method as the lower bound on the true difference for the hypothesis test). The proportion of the bootstrap distribution below zero provides a bootstrap percentile p-value for the hypothesis of no improvement versus a positive improvement in capture threshold.

Due to the slight skewness in the shape of the histogram that can be seen in Figures 3.1 and 3.3, the Pacesetter statisticians were concerned that the percentile method for determining the bootstrap lower confidence bound on the difference in the mean values might not be sufficiently accurate.

The bootstrap percentile t method was considered, but time did not permit the method to be developed in time for the submission. In a later clinical trial, Pacesetter took the same approach with the comparison of the control and treatment for the Passive Plus DX clinical trial.

The bootstrap percentile t method is a simple method to program and appears to overcome some of the shortcomings of Efron's percentile method without the complications of bias correction and acceleration constants. This technique was first presented by Efron as the bootstrap (Efron, 1982a, Section 10.10). Later, in Hall (1986a) asymptotic formulas were developed for the coverage error of the bootstrap percentile t method. This is the method discussed previously in Section 3.1.5.

The Passive Plus DX lead is a passive fixation steroid eluting lead that was compared with a non-steroid approved version of the lead. The 3:1 randomization of treatment group to control group was used in the Passive Plus study also.

In the Passive Plus study, the capture thresholds behaved similarly to those for the leads in the Tendril DX study. The main difference in the results was that the mean differences were not quite as large (i.e., close to the 0.5-volt improvement for the steroid lead over the non-steroid lead, whereas for Tendril DX the improvement was close to a 1.0-volt improvement).

In the Passive Plus study, both the bootstrap percentile method lower 95% and the bootstrap percentile t method lower 95% confidence bounds were determined.

3.4. AN APPLICATION OF BOOTSTRAP CONFIDENCE INTERVALS TO BINARY DOSE–RESPONSE MODELING

At pharmaceutical companies, a major part of the early phase II development is the establishment of a dose–response relationship for a drug that is being considered for marketing. At the same time estimation of doses that are minimally effective or maximally safe are important to determine what is the best dose or small set of doses to carry over into phase III trials. The following

example, Klingenberg (2007), was chosen because it addresses methods that are important in improving the phase 2 development process for new pharmaceuticals (an important application) and it provides an example where resampling methods are used in a routine fashion. Permutation methods are used for p-value adjustment due to multiplicity, and bootstrap confidence intervals are used to estimate the minimum effective dose after proof of concept.

In the spirit of faster development of drugs through adaptive design concepts, Klingenberg (2007) proposes a unified approach to determining proof of concept with a new drug followed by dose–response modeling and dose estimation. In this paper, Klingenberg describes some of the issues that have motivated this new statistical research. The purpose of the paper is to provide a unified approach to proof of concept (PoC) phase 2a clinical trials with the dose finding phase 2b trials in an efficient way when the responses are binary. The goal at the end of phase 2 is to find a dose for the drug that will be safe and effective and therefore will have a good chance for success in phase 3. Klingenberg cites the following statistics as an indication of the need to find different approaches that have better chances of achieving the phase 2 objectives.

He notes that the current failure rate for phase 3 trials is approaching 50%, largely attributed to improper target dose estimation/selection in phase II and incorrect or incomplete knowledge of the dose–response, and the FDA reports that 20% of the approved drugs between 1980 and 1989 had the initial dose changed by *more than 33%*, in most cases lowering it. So current approaches to phase 2 trials are doing a poor job of achieving the objectives since poor identification of dose is leading to the use of improper doses that lead to wasted phase 3 trials and even when the trials succeed, they often do so with a less than ideal choice of dose and in the post-marketing phase the dose is determined to be too high and reduced dramatically.

The idea of the approach is to use the following strategy: (1) Work with the clinical team to identify a reasonable class of potential dose–response models; (2) from this comprehensive set of models, choose the ones that best describe the dose–response data; (3) use model averaging to estimate a target dose; (4) decide which models, if any, significantly pick up the signal, establishing PoC; (5) use the permutation distribution of the maximum penalized deviance over the candidate set to determine the best model (s0); and (6) use the best model to estimate the minimum effective dose (MED). Important aspects of the approach are the use of permutation methods to determine adjusted p-values and control the error rate of declaring spurious signals as significant (due to the multiplicity of models considered). A thorough evaluation and comparison of the approach to popular contrast tests reveals that its power is as good or better in detecting a dose–response signal under a variety of situations, with many more additional benefits: It incorporates model uncertainty in proof of concept decisions and target dose estimation,

yields confidence intervals for target dose estimates (MED), allows for adjustments due to covariates, and extends to more complicated data structures. Klingenberg illustrates his method with the analysis of a Phase II clinical trial.

The bootstrap enters into this process as the procedure for determining confidence intervals for the dose. Permutation methods due to Westfall and Young (1993) were used for the p-value adjustment. Westfall and Young (1993) also devised a bootstrap method for p-value adjustment that is very similar to the permutation approach and could also have been used. We cover the bootstrap method for p-value adjustment with some applications in Chapter 8.

The unified approach that is used by Klingenberg is similar to the approach taken by Bretz, Pinheiro, and Branson (2005) for normally distributed data but applied to binomial distributed data. MED estimation in this paper follows closely the approach of Bretz, Pinheiro, and Branson (2005). A bootstrap percentile method confidence interval for MED is constructed using the fit to the chosen dose-response model. The confidence interval is constructed conditional on the establishment of PoC. Klingenberg illustrates the methodology by reanalyzing data from a phase 2 clinical trial using a unified approach.

In Klingenberg's example, the key variable is a binary indicator for the relief of symptoms from irritable bowl syndrome (IBS), a disorder that is reported to affect up to 30% of all Americans at sometime during their lives (American Society of Colon and Rectal Surgeons, www.fascrs.org). A phase II clinical trial investigated the effcacy of a compound against IBS in women at $k = 5$ dose levels ranging from placebo to 24 mg. Expert opinion was used to determine a target dose. Here, Klingenberg reanalyzes these data within the statistical framework of the unified approach.

Preliminary studies with only two doses indicated a placebo effect of roughly 30% and a maximal possible dose effect of 35%. However, prior to the trial, investigators were uncertain about the monotonicity and curvature of a possible dose effect. The first eight models and the zero effect model are pictured in Figure 3.5 for a particular prior choice of parameter values, cover a broad range of dose–response shapes deemed plausible for his particular compound, and were selected to form the candidate set. The candidate models had to be somewhat broad because the investigators could not rule out strongly concave or convex patterns or even a down-turn at higher doses, and hence the candidate set includes models to the see these possible effects. All models in Figure 3.5, most with fractional polynomial (Roystone and Altman, 1994) linear predictor form, are fit to the data by maximum likelihood, but some of the models might not converge for every possible data set.

The author is interested in models that pick up a potential signal observed in a dose–response study. To this end, he compared each of the eight models to the model of no dose effect via a (penalized) likelihood ratio test. A description of the models is given in Table 3.2.

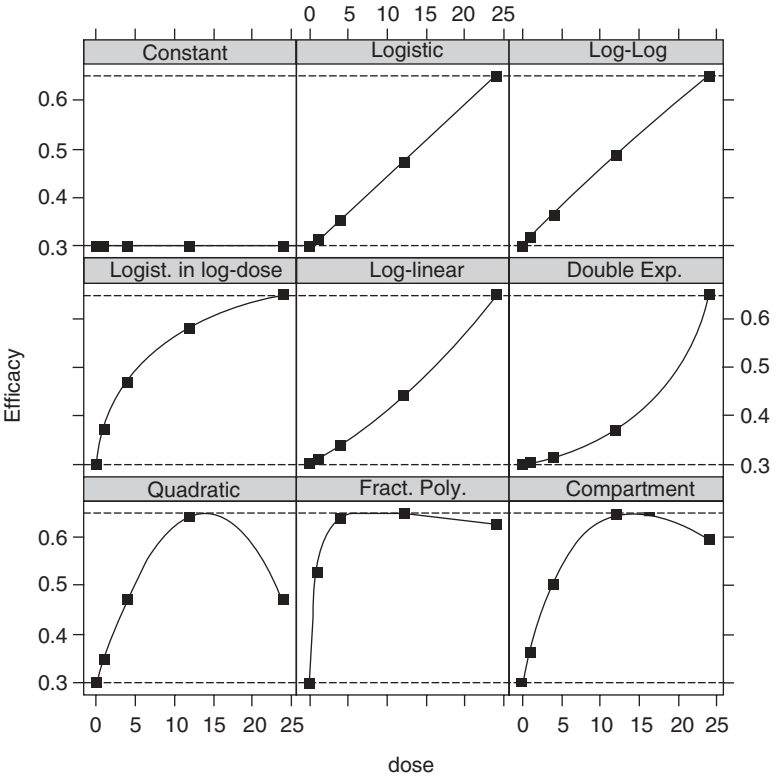


Figure 3.5 A zero effect model and eight candidate dose–response models. [Taken from Klingenberg (2007) with permission.]

Table 3.2 Dose–Response Models for the Efficacy of Irritable Bowel Syndrome Compound

M	Model	Link	Predictor	Number of Permutations
M_1:	Logistic	logit	$\beta_0 + \beta_1 d$	2
M_2:	Log-Log	log–log	$\beta_0 + \beta_1 d$	2
M_3:	Logistic in log-dose	logit	$\beta_0 + \beta_1 \log(d + 1)$	2
M_4:	Log-linear	log	$\beta_0 + \beta_1 d$	2
M_5:	Double-exponential	identity	$\beta_0 + \beta_1 \exp(\exp(d/\max(d)))$	2
M_6:	Quadratic	identity	$\beta_0 + \beta_1 d + \beta_2 d^2$	3
M_7:	Fractional Poly	logit	$\beta_0 + \beta_1 \log(d + 1) + \beta_2/(d + 1)$	3
M_8:	Compartment	identity	$\beta_0 + \beta_1 d \exp(-d/\beta_2)$, $\beta_2 > 0$	3
M_9:	Square root	logit	$\beta_0 + \beta_1 d^{1/2}$	2
M_{10}:	Emax	logit	$\beta_0 + \beta_1 d/(\beta_2 + d)$, $\beta_2 > 0$	3

Source: Taken from Klingenberg (2007) with permission.

Table 3.3 G_s^2-Statistics, p-Values, Target Dose Estimates and Model Weights

Model Number	Model Type	G_s^2	Raw p-Value	Adjusted p-Value	MED (mg)	Model Weight (%)
M_1	Logistic	3.68	0.017	0.026	N/A	0
M_2	Log–Log	3.85	0.015	0.024	N/A	0
M_3	Logistic in log-dose	10.53	$<10^{-3}$	0.001	7.9	6
M_4	Log-linear	3.25	0.022	0.032	N/A	0
M_5	Double-exponential	0.90	0.088	0.106	N/A	0
M_6	Quadratic	6.71	0.005	0.005	7.3	1
M_7	Fractional Polynomial	15.63	$<10^{-4}$	$<10^{-4}$	0.7	81
M_8	Compartment	11.79	$<10^{-3}$	$<10^{-3}$	2.5	12
		Critical value 2.40			MED (avg.) = 1.4	
					95% Conf. Int. = [0.4, 12.0]	

Source: Adapted from Klingenberg (2007) with permission.

Table 3.3 gives the results for the tests of the models including the raw and adjusted p-values. Also included are the weights used in the model averaging. For each model that was included the point estimate of the MED is given. Also we see that the weighted average of the four selected models is 1.4 and the 95% bootstrap percentile confidence interval is [0.4, 12.0]. The critical value for the null (permutation) distribution of the maximum penalized deviance is shown to be 2.4, and seven of the eight models (all but M_5) have a test statistic that exceeds the critical value. But only models (M_3, M_6, M_7, and M_8) were used in the final averaging.

3.5. HISTORICAL NOTES

Bootstrap confidence intervals were introduced in Efron (1982a, Chapter 10). Efron's percentile method and the bias corrected percentile method were introduced at that time. Efron also introduced in Efron (1982a) the bootstrap t intervals and illustrated these techniques with the median as the parameter.

It was recognized at that time that confidence interval estimation was a tougher problem than estimating standard errors and considerably more bootstrap samples would be required (i.e., 1000 bootstrap samples for confidence intervals where only 100 would be required for standard error estimates). For more discussion of this issue see Section 7.1.

In an important work, Schenker (1985) shows that bias adjustment to Efron's percentile method is not always sufficient to provide "good" confi-

dence intervals. Nat Schenker's examples motivated Efron to come up with the use of an acceleration constant as well as a bias correction in the modification of the confidence interval endpoints. This led to a significant improvement in the bootstrap confidence intervals and removed Schenker's objections.

The idea of bootstrap iteration to improve confidence interval estimation appears in Hall (1986a), Beran (1987), Loh (1987), Hall and Martin (1988a), and DiCiccio and Romano (1988). The methods of Hall, Beran, and Loh all differ in the way they correct the critical point(s). Loh refers to his approach as bootstrap calibration.

Hall (1986b) deals with sample size requirements. Specific application to the confidence interval estimation for the correlation coefficient is given in Hall, Martin, and Schucany (1989). For further developments in bootstrap iteration see Martin (1990a), Hall (1992a), or Davison and Hinkley (1997).

Some of the asymptotic theory is based on formal Edgeworth expansions that were rigorously developed in Bhattacharya and Ghosh (1978) [see Hall (1992a) for a detailed account with applications to the bootstrap]. Other asymptotic expansions such as saddlepoint approximations may provide comparable confidence intervals without the need for Monte Carlo [see the monograph by Field and Ronchetti (1990) and the papers by Davison and Hinkley (1988) and Tingley and Field (1990)].

DiCiccio and Efron (1992) also obtain very good confidence intervals without Monte Carlo for data from an exponential family of distributions. DiCiccio and Romano (1989a) also produce accurate confidence limits by making some parametric assumptions.

Some the research in the 1980s and late 1990s suggests that the Monte Carlo approximation may not be necessary (see Section 7.3 and the references above) or that the number of Monte Carlo replications can be considerably reduced by variance reduction techniques [see Section 7.2 and Davison, Hinkley, and Schechtman (1986), Therneau (1983), Hesterberg (1988), Johns (1988), and Hinkley and Shi 1989]. The most recent developments can be found in Hesterberg (1995a,b, 1996, 1997).

Discussions of bootstrap hypothesis tests appear in the early paper of Efron (1979a) and some work can be found in Beran (1988c), Hinkley (1988), Fisher and Hall (1990) and Hall and Wilson (1991). Specific applications and Monte Carlo studies of bootstrap hypothesis testing problems are given in Dielman and Pfaffenberger (1988), Rayner (1990a,b), and Rayner and Dielman (1990).

Fisher and Hall (1990) point out that even though there are close connections between bootstrap hypothesis tests and confidence intervals there are also important differences which lead to specialized treatment. They recommend the use of asymptotic pivotal quantities in order to maintain a close approximation to the advertised significance level for the test.

Ideas are illustrated using the analysis of variance problem with both real and simulated data sets. Results based on Edgeworth expansions and Cornish–Fisher expansions clearly demonstrate the advantage of bootstrapping pivotal

statistics for both hypothesis testing and confidence intervals [see Hall (1992a)]. Lehmann (1986) is the second edition of a classic reference on hypothesis testing and any reader wanting a rigorous treatment of the subject would be well advised to consult that text.

The first application of Edgeworth expansions to derive properties for the bootstrap is Singh (1981). The work of Bickel and Freedman (1981) is similar to that of Singh (1981) and also uses Edgeworth expansions. Their work shows how bootstrap methods correct for skewness.

Both papers applied one-term Edgeworth expansion corrections. Much of the development of Edgeworth expansions goes back to the determination of particular cumulants, as in James (1955, 1958).

The importance of asymptotically pivotal quantities was not brought out in the early papers because the authors considered a nonstudentized sample mean and assumed the population variance is known. Rather this result was first mentioned by Babu, and Singh in a series of papers (Babu and Singh, 1983, 1984a, and 1985). Another key paper on the use of Edgeworth expansions for hypothesis testing is Abramovitch and Singh (1985).

Hall (1986a, 1988b) wrote two key papers which demonstrate the value of asymptotically pivotal quantities in the accuracy of bootstrap confidence intervals.

Hall (1986a) derives asymptotic formulas for coverage error of the bootstrap percentile t confidence intervals and Hall (1988b) gives a general theory for bootstrap confidence intervals. Theoretical comparisons of variations on bootstrap percentile t confidence intervals are given in Bickel (1992). Other papers that support the use of pivotal statistics are Beran (1987) and Liu and Singh (1987).

Methods based on symmetric bootstrap confidence intervals are introduced in Hall (1988a). Hall also defines "short" bootstrap confidence intervals in Hall (1988b) [see also Hall (1992a) for some discussions]. The idea for the "short" bootstrap confidence intervals goes back to Buckland (1980, 1983).

Efron first proposed his version of the percentile method in Efron (1979a) [see also Efron (1982a) for detailed discussions]. The BC_a intervals were first given in Efron (1987). Buckland (1983, 1984, 1985) provide applications for Efron's bias correction intervals along with algorithms for their construction.

Bootstrap iteration in the context of confidence intervals is introduced in Hall (1986a) and Beran (1987). Hall and Martin (1988a) develop a general framework for bootstrap iteration. Loh (1987) introduced the notion of bootstrap calibration. When applied to bootstrap confidence intervals, calibration is equivalent to bootstrap iteration.

Other important works related to confidence intervals and hypothesis testing include Beran (1986, 1990a,b).

CHAPTER 4

Regression Analysis

This chapter is divided into three parts and a historical notes section. Section 4.1 deals with linear regression and Section 4.2 deals with the nonlinear regression problems. Section 4.3 deals with nonparametric regression models. In Section 4.4 we provide historical notes regarding the development of the bootstrap procedures in both the linear and nonlinear cases.

In Section 4.1.1 we will briefly review the well-known Gauss–Markov theory, which applies to least-squares estimation in the linear regression problem. A natural question for the practitioner is to ask "Why bootstrap in the linear regression case? Isn't least-squares a well-established approach that has served us well in countless applications?" The answer is that for many problems, least-squares regression has served us well and is always useful as a first approach but is problematic when the residuals have heavy-tailed distributions or if even just a few outliers are present.

The difficulty is that in some applications, certain key assumptions may be violated. These assumptions are as follows: (1) The error term in the model has a probability distribution that is the same for each observation and does not depend on the predictor variables (i.e., independence and homoscedasticity); (2) the predictor variables are observed without error; and (3) the error term has a finite variance.

Under these three assumptions, the least-squares procedure provides the best linear unbiased estimate of the regression parameters. However, if assumption 1 is violated because the variance of the residuals varies as the predictor variables change, a weighted least-squares approach may be more appropriate.

The strongest case for least-squares estimation can be made when the error term has a Gaussian or approximately a Gaussian distribution. Then the theory of maximum likelihood also applies and confidence intervals and

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

hypothesis tests for the parameters can be applied using the standard theory and the standard statistical packages.

However, if the error distribution is non-Gaussian and particularly if the error distribution is heavy-tailed, least-squares estimation may not be suitable (robust regression methods may be better). When the error distribution is non-Gaussian, regardless of what estimation procedure is used, it is difficult to determine confidence intervals for the parameters or to obtain prediction intervals for the response variable.

This is where the bootstrap can help, and we will illustrate it for both the linear and nonlinear cases. In the nonlinear case, even standard errors for the estimates are not easily obtained, but bootstrap estimates are fairly straightforward.

There are two basic approaches to bootstrapping in the regression problem. One is to first fit the model and bootstrap the residuals. The other is to bootstrap the vector of the response variables and the associated predictor variable. Bootstrapping the residuals requires that the residuals be independent and identically distributed (or at least exchangeable).

In a quasi-optical experiment (Shimabukuro, Lazar, Dyson, and Chernick, 1984), I used the bootstrap to estimate the standard errors for two of the parameters in the nonlinear regression model. Results are discussed in Section 4.2.2. The residuals appear to be correlated with the incident angle of the measurement. This invalidates the exchangeability assumption, but how does it affect the standard errors of the parameters?

Our suspicion is that bootstrapping the residuals makes the bootstrap sample more variable and consequently biases the estimated standard errors on the high side. This, however, remains an open question. Clearly, from the intuitive point of view the bootstrapping is not properly mimicking the variation in the actual residuals and the procedure can be brought into question.

A second method with more general applicability is to bootstrap the vector of the observed response variable and the associated predictor variables. This only requires that the vectors are exchangeable and does not place explicit requirements on the residuals from the model.

However, some statisticians, particularly from the British school, view the second method philosophically as an inappropriate approach. To them, the regression problem requires that the predictor variables be fixed for the experiment and not selected at random from a probability distribution. The bootstrapping of the vector of response and predictor variables implicitly assumes a joint probability distribution for the vector of predictor variables and response. From their point of view, this is an inappropriate model and hence the vector approach is not an option.

However, from the practical point of view, if the approach of bootstrapping the vector has nice robustness properties related to model specification, it is justified. This was suggested by Efron and Tibshirani (1993, p. 113) for the case of a single predictor variable. Since it is robust, it is not important whether or not the method closely mimics the assumed but not necessarily correct

regression model. Presumably their observation extends to the case of more than one predictor variable.

On the other hand, some might argue that bootstrapping the residuals is only appropriate when the predictor variables are not fixed. This comes down to another philosophical issue that only statisticians care about. The question is one of whether conditional inference is valid when the experiment really involves an unconditional joint distribution for the predictor and response variables.

This is a familiar technical debate for statisticians because it is the same issue regarding the appropriateness of conditioning on the marginal totals in a 2×2 contingency table. Conditioning on ancillary information in the data (i.e., information in the data that does not have any affect on the "best" estimate of a parameter is a principle used by Sir Ronald Fisher in his theory of inference and is best known to be applied in Fisher's exact permutation test, which is most commonly used in applications involving categorical data).

For the practitioner, I repeat the sage advice of my friend and former colleague, V. K. Murthy, who often said "the proof of the pudding is in the eating." This applies here to these bootstrap regression methods as it does in the comparison of variants of the bootstrap in discriminant analysis. If we simulate the process under accepted modeling assumptions, the method that performs best in the simulation is the one to use regardless of how much you believe or like some particular theory.

These two methods for bootstrapping in regression are given by Efron (1982a, pp. 35–36). These methods are very general. They apply to linear and nonlinear regression models and can be used for least-squares or for any other estimation procedure. We shall now describe these bootstrap methods.

A general regression model can be given by

$$Y_i = g_i(\boldsymbol{\beta}) + \varepsilon_i \qquad \text{for } i = 1, 2, \ldots, n.$$

The functions g_i are of known form and may depend on a fixed vector of covariates c_i. The vector $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and the ε_i are independent and identically distributeed with some distribution F.

We assume that F is "centered" at zero. Usually this means that the expected or average value of ε_i is zero. However, in cases where the expected value does not exist, we may use the criterion that $P(\varepsilon < 0) = 0.50$.

Given the observed vector

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

where the ith component y_i is the observed value of the random variable Y_i, we find the estimate of $\boldsymbol{\beta}$, which minimizes the distance measure between $\boldsymbol{y}$ and $\boldsymbol{\lambda}(\boldsymbol{\beta})$ where

$$\boldsymbol{\lambda}(\boldsymbol{\beta}) = \begin{pmatrix} g_1(\beta) \\ g_2(\beta) \\ \vdots \\ g_n(\beta) \end{pmatrix}.$$

Denote the distance measure by $D(\boldsymbol{y},\boldsymbol{\lambda}, (\boldsymbol{\beta}))$. If

$$\mathbf{D}(\boldsymbol{y}, \boldsymbol{\lambda}, (\boldsymbol{\beta})) = \textstyle\sum_1^n [(y_i - g_i(\beta)]^2,$$

we get the usual least-squares estimates. For least absolute deviations, we would choose

$$\mathbf{D}(\boldsymbol{y}, \boldsymbol{\lambda}, (\boldsymbol{\beta})) = \textstyle\sum_1^n |[y_i - g_i(\beta)]|.$$

Now by taking $\hat{\beta} = \min_{\beta} \mathbf{D}(\boldsymbol{y}, \boldsymbol{\lambda}, (\boldsymbol{\beta}))$, we have our parameter estimate of $\boldsymbol{\beta}$. The residuals are then obtained as $\hat{\varepsilon}_i = y_i - g_i(\hat{\beta})$.

The first bootstrap approach is to simply bootstrap the residuals. This is accomplished by constructing the distribution F_n that places probability $1/n$ at each $\hat{\varepsilon}_i$ We then generate bootstrap residuals ε_i^* for $i = 1, 2, \ldots, n$, where the ε_i^* are obtained by sampling independently from F_n (i.e., we sample with replacement from $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$). We then have a bootstrap sample data set;

$$y_i^* = g_i(\hat{\beta}) + \varepsilon_i^* \qquad \text{for } i = 1, 2, \ldots, n.$$

For each such bootstrap data set y^*, we obtain

$$\hat{\beta}^* = \min_{\beta}[y^*, \lambda, \beta].$$

The procedure is repeated B times and the covariance matrix for $\hat{\beta}$ is estimated as t, where $\hat{\beta}_j^*$ is the bootstrap estimate from the jth bootstrap sample and $\hat{\beta}_*^* = \frac{1}{B}\sum_{j=1}^{B} \hat{\beta}_j^*$. This is the covariance estimate suggested by Efron (1982a, p. 36).

We note that bootstrap theory suggests simply using $\hat{\beta}$ in place of $\hat{\beta}_*^*$. The resulting covariance estimate should be close to that suggested by Efron. Confidence intervals for β can be obtained by the methods described in Chapter 3, but with the bootstrap samples for the $\hat{\beta}$ values.

The second approach is to bootstrap the vector

$$Z = \begin{pmatrix} y_i \\ c_i \end{pmatrix}$$

of the observations y_i and the covariates or predictor variables $\boldsymbol{c}_i$ for $i = 1, 2, \ldots, n$. The bootstrap samples are then z_i^* for $i = 1, 2, \ldots, n$ obtained by giving probability of selection $1/n$ to each z_i. Taking $z_i^* = \begin{pmatrix} y_i^* \\ c_i^* \end{pmatrix}$, we use y_i^* to obtain the $\hat{\beta}^*$ just as before.

Efron claims that although the two approaches are asymptotically equivalent for the given model, the second approach is less sensitive to model misspecification. It also appears that since we do not bootstrap the residuals, the second approach may be less sensitive to the assumptions concerning independence or exchangeability of the error terms.

4.1. LINEAR MODELS

In the case of the linear regression model, if the least-squares estimation procedure is used, there is nothing to be gained by bootstrapping. As long as the error terms are independent and identically distributed with mean zero and common variance σ^2, the least-squares estimates of the regression parameters will be the best among all linear unbiased estimators. The covariance matrix corresponding to the least-squares estimate $\hat{\beta}$ of the parameter vector $\boldsymbol{\beta}$ is given by

$$\sum = \sigma^2 (X^T X)^{-1},$$

where X is called the design matrix and $(X^T X)^{-1}$ is well-defined if X is a full-rank matrix. If $\hat{\sigma}^2$ is the least-squares estimate of the residual variance σ^2, then

$$\hat{\sum} = \hat{\sigma}^2 (X^T X)^{-1}$$

is the commonly used estimate of the parameter covariance matrix.

For more details see Draper and Smith (1981). These least-squares estimates are the standard estimates that can be found in all the standard statistical computer programs.

If, in addition, the error terms are Gaussian or approximately Gaussian, the least-squares estimates are also the maximum likelihood estimates. Also, the confidence intervals for the regression parameters, hypotheses tests about the parameters, and prediction intervals for a new observation based on known values of the regression variables can be determined in a straightforward way.

In the non-Gaussian case, even though we can estimate the parameter covariance matrix, we will not know the probability distribution for $\hat{\beta}$ and so we cannot determine confidence intervals and prediction intervals or perform hypothesis tests using the standard methods. The bootstrap approach does, however, provide a method for approximating the distribution of $\hat{\beta}$ through bootstrap sample estimates $\hat{\beta}^*$.

First we review the Gauss–Markov theory of least-squares estimation in Section 4.1.1. In Section 4.1.2 we discuss, in more detail, situations where we might prefer to use other estimates of β such as the least absolute deviation estimates or *M*-estimates.

In Section 4.1.3 we discuss bootstrap residuals and the possible problems that can arise. If we bootstrap the vector of response and predictor variables, we can avoid some of the problems of bootstrapping residuals.

4.1.1. Gauss–Markov Theory

The least-squares estimator of the regression parameters are maximum likelihood when the error terms is assumed to be Gaussian. Consequently, the least-squares estimates have the usual optimal properties under the Gaussian model. They are unbiased and asymptotically efficient. In fact, they have the minimum variance among unbiased estimators.

The Gauss–Markov theorem is a more general result in that it applies to linear regression models with general error distributions. All that is assumed is that the error distribution has mean zero and variance σ^2. The theorem states that among all estimators that are both unbiased and a linear function of the responses y_i for $i = 1, 2, \ldots, n$ the least-squares estimate has the smallest possible variance.

The result was first shown by Carl Friedrich Gauss in 1821. For more details about the theory, see the *Encyclopedia of Statistical Science*, Vol. 3, pp. 314–316.

4.1.2. Why Not Just Use Least Squares?

In the face of all these optimal properties, one should ask why least squares shouldn't always be the method of choice? The basic answer is that the least-squares estimates are very sensitive to violations in the modeling assumptions. If the error distribution has heavy tails or the data contain a few "outliers," the least-squares estimates will not be very good.

This is particularly true if these outliers are located at high leverage points (i.e., points that will have a large influence on the slope parameters). High leverage points occur at or near the extreme values of the predictor variables. In cases of heavy tails or outliers, the method of least absolute deviations or other robust regression procedures such as *M*-estimation or the method of repeated medians provide better solutions though analytically they are more complex.

Regardless of the procedure used, we may be interested in confidence regions for the regression parameters or prediction intervals for future cases. Under the Gaussian theory for least squares, this is possible. However, if the error distribution is non-Gaussian and unknown, the bootstrap provides a method for computing standard errors for the regression parameters or prediction intervals for future values, regardless of the method of estimation.

There are many other complications to the regression problem that can be handled by bootstrapping. These include the problem of heteroscedasticity of the variance of the error term, nonlinearity in the model terms, and bias adjustment when transformation of variables is used.

For a bootstrap-type approach to the problem of retransformation bias, see Duan (1983). Bootstrap approaches to the problem of heteroscedasticity are covered in Carroll and Ruppert (1988).

An application of bootstrapping residuals for a nonlinear regression problem is given in Shimbukuro et al. (1984) and will be discussed later. When procedures other than least-squares are used, confidence intervals and prediction intervals are still available by bootstrapping.

Both editions of a book by Miller (1986, 1997) deal with linear models. These are very excellent references for the understanding of the importance of modeling assumptions. They also demonstrate when and why the methods are robust to departures from basic assumptions. These texts also point out when robust and bootstrap statistical procedures are more appropriate.

4.1.3. Should I Bootstrap the Residuals from the Fit?

From Efron (1979a, Section 7), the bootstrap estimate of the covariance matrix for the coefficients in a linear regression model is shown to be

$$\hat{\sum} = \sigma^2 \left(\sum_{i-1}^{\hat{a}} c_i c_i \right)^{-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{a} \hat{\varepsilon}_i^2.$$

The model is given by $y_i = c_i \boldsymbol{\beta} + \varepsilon_i$ for $i = 1, 2, \ldots, n$ and $\hat{\varepsilon}_i$ is the residual estimate obtained by least-squares. The only difference between this estimate and the standard one from the Gauss–Markov theory is use of n in the denominator of the estimate for σ^2. The standard theory would use $n - p$, where p is the number of the covariates in the model (i.e., the dimension of the vector

β). So we see that at least when the linear least-squares model is an appropriate method bootstrapping the residuals gives nearly the same answer as the Gauss–Markov theory for larger. Of course, in such a case, we do not need to bootstrap since we already have an adequate model.

It is important to ask how well this approach to bootstrapping residuals works when there is not an adequate theory for estimating the covariance matrix for the regression parameters. There are many situations that we would like to consider: (1) heteroscedasticity in the residual variance; (2) correlation structure in the residuals; (3) nonlinear models; (4) non-Gaussian error distributions; and (5) more complex econometric and time series models.

Unfortunately, the theory has not quite reached the level of maturity to give complete answers in these cases. There are still many open research questions to be answered. In this section and in Section 4.2, we will try to give partial answers to (1) through (4); (5) is being deferred to Chapter 5, which covers time series methods.

A second approach to bootstrapping in a regression problem is to bootstrap the entire vector

$$Z_i = \begin{pmatrix} y_i \\ c_i \end{pmatrix}$$

that is a $(p + 1)$-dimensional vector of the response variable and the covariate values. A bootstrap sample is obtained by choosing integers at random with replacement from the set 1, 2, 3, . . . , n until n integers have been chosen. If, on the first selection, say, integer j is chosen, then the bootstrap observation is $Z_i^* = Z_j$. After a bootstrap sample has been chosen, the regression model is fit to the bootstrap samples producing an estimate β^*. By repeating this B times, we get $\beta_1^*, \beta_2^*, \ldots, \beta_B^*$ the bootstrap sample estimates of β. The usual sample estimates of variance and covariance can then be applied to $\beta_1^*, \beta_2^*, \ldots, \beta_B^*$.

Efron and Tibshirani (1986) claim that the two approaches are asymptotically equivalent (presumably when the covariates are assumed to be chosen from a probability distribution), but can perform differently in small sample situations.

The latter method does not take full advantage of the special structure of the regression problem. Whereas bootstrapping the residuals leads to the estimates $\hat{\Sigma}$ and $\hat{\sigma}^2$ as defined earlier when $B \to \infty$, this latter procedure does not.

The advantage is that it provides better estimates of the variability in the regression parameters when the model is not correct. We recommend it over bootstrapping the residuals when (1) there is heteroscedasticity in the residual variance, (2) there is correlation structure in the residuals, or (3) we

suspect that there may be other important parameters missing from the model.

Wu (1986) discusses the use of a jackknife approach in regression analysis which he views to be superior to the bootstrap approaches we have mentioned. His approach works particularly well in the case of heteroscedasticity of residual variances.

There are several discussants to Wu's paper. Some strongly support the bootstrap approach and point out modifications for heteroscedastic models, Wu claims that even such modifications to the bootstrap will not work for nonlinear and binary regression problems. The issues are far from settled.

The two bootstrap methods described in this section apply equally to nonlinear (homoscedastic, i.e., constant variance) models as well as the linear (homoscedastic) models. In the next section, we will give some examples of nonlinear models. We will then consider a particular experiment where we bootstrap the residuals.

4.2. NONLINEAR MODELS

The theory of nonlinear regression models has advanced greatly in the 1970s and 1980s. Much of this development has been well-documented in recent textbooks devoted strictly to nonlinear models. Two such books are Bates and Watts (1988) and Gallant (1987).

The nonlinear models can be broken up into two categories. In the first category, local linear approximations can be made using Taylor series, for example. When this can be done, approximate confidence or prediction intervals can be generated based on asymptotic theory.

Much of this theory is covered in Gallant (1987). In the aerospace industry, there has been great success applying local linearization methods in the construction of Kalman filters for missiles, satellites and other orbiting objects.

The second category is the highly nonlinear model for which the linear approximation will not work. Bates and Watts (1988) provide methods for diagnosing the severity of the nonlinearity.

The bootstrap method can be applied to any type or nonlinear model. The two methods as described in Efron (1982a) can be applied to fairly general problems. To bootstrap, we do not need to have a differentiable functional form. The nonlinear model could even be a computer algorithm rather than an analytical expression. We do not need to restrict the residual variance to have a Gaussian distribution. The only requirements are that the residuals should be independent and identically distributed (exchangeable may be sufficient) and their distribution should have a finite variance. The distribution of the residuals should not change as the predictor variables are changed. This requirement imposes homoscedasticity on the residual variance.

The distribution of the residuals should not change because the predictor variables changed. This requirement imposes homoscedasticity on the residual variance.

For models with heteroscedastic variance, modifications to the bootstrap are available. We shall not discuss these modifications here. To learn more about it, look at the discussion to Wu (1986).

4.2.1. Examples of Nonlinear Models

In Section 4.2.2, we discuss a quasi-optical experiment that was performed to determine the accuracy of a new measurement technique for the estimation of optical properties of materials used to transmit and/or receive millimeter wavelength signals. This experiment was conducted at the Aerospace Laboratory.

As a statistician in the engineering group, I was asked to determine the standard errors of their estimates. The statistical model was nonlinear and I chose to use the bootstrap to estimate the standard error. Details on the model and the results of the analysis are given in Section 4.2.2.

Many problems that arise in practice can be solved by approximate models that are linear in the parameters (remember that in statistical models the distinction between linear and nonlinear is in the parameters and not in the predictor variables). The scope of applicability of linear models can, at times, be extended by including transformations of the variables.

However, there are limits to what can adequately be approximated by linear models. In many practical scientific endeavors, the model may arise from a solution to a differential equation. A nonlinear model that could arise as the solution of a simple differential equation might be the function

$$f(x, \sigma)=\sigma_1 + \sigma_2 \exp(\sigma_3 x),$$

where x is a predictor variable and

$$\sigma = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix}$$

is a three-dimensional parameter vector.

A common problem in time series analysis is the so-called harmonic regression problem. We may know that the response function is periodic or the sum of a few periodic functions, but we do not know the amplitude or the frequency of the periodic components. Here it is the fact that the frequencies are among the unknown parameters that makes the model nonlinear. The simple case of a single periodic function can be described by the following function.

$$f(t, \varphi) = \varphi_0 + \varphi_1 \sin(\varphi_2 t + \varphi_3)$$

where t is the time since a specific epoch and

$$\varphi = \begin{pmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix}$$

is a vector of unknown parameters. The parameters φ_1, φ_2, and φ_3 all have physical interpretations. φ_1 is called the amplitude, φ_2 is the frequency, and φ_3 is the phase delay.

Because of the trigonometric identity

$$\sin(A + B) = \sin A \cos B + \cos A \sin B$$

we can reexpress

$$\varphi_1 \sin(\varphi_2 t \varphi_3)$$

as

$$\varphi_1 \cos \varphi_3 \sin \varphi_2 t + \varphi_1 \sin \varphi_3 \cos \varphi_2 t.$$

The problem can then be reparameterized as

$$f(t, A) = A_0 + A_1 \sin A_2 t + A_3 \cos A_2 t,$$

where

$$A = \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix}$$

and $A_0 = \varphi_0$, $A_1 = \varphi_1 \cos\varphi_3$, $A_2 = \varphi_2$, and $A_3 = \varphi_1 \sin \varphi_3$.

This reparameterized form of the model is the form given by Gallant (1987, p. 3) with slightly different notation.

There are many other examples where nonlinear models are solutions to differential equations or systems of differential equations. Even in the case of linear differential equations or systems of linear differential equations, the

solutions involve exponential functions (both real- and complex-valued). The results are then real-valued functions that are periodic or exponential or a combination of both.

If constants involved in the differential equation are unknown, then their estimates will be obtained through the solution of a nonlinear model. As a simple example, consider the equation

$$\frac{d}{dx}y(x) = -\sigma_1 y(x)$$

subject to the initial condition y(0) = 1. The solution is then

$$y(x) = e^{-\varphi_1 x}.$$

Since φ_1 is an unknown parameter, the function $y(x)$ is nonlinear in φ_1.

For a commonly used linear system of differential equations whose solution involves a nonlinear model, see Gallant (1987, pp. 5–8). Such systems of differential equations arise in compartmental analysis commonly used in chemical kinetics problems.

4.2.2. A Quasi-optical Experiment

In this experiment, I was asked as a consulting statisticia n to determine estimates of two parameters that were of interest to the experimenters. More importantly, they needed a "good" estimate of the standard errors of these estimates since they were proposing a new measurement technique that they believed would be more accurate than previous methods.

Since the model was nonlinear and I was given a computer program rather than an analytic expression, I chose to bootstrap the residuals. The results were published in Shimabukuro, Lazar, Dyson, and Chernick (1984).

The experimenters were interested in the relative permittivity and the loss tangent (two material properties related to the transmission of signals at millimeter wavelengths through a dielectric slab). The experimental setup is graphically depicted in Figure 4.1. Measurements are taken to compute $|T|^2$, where T is a complex number called the transmission coefficient. An expression for T is given by

$$T = \frac{(1-r^2)e^{-(\beta_1-\beta_0)di}}{1-r^2e^{-2\beta_1 di}},$$

where

$$\beta_1 = \frac{2\pi}{\lambda_0}\sqrt{\varepsilon_1/\varepsilon_0 - \sin^2\varphi},$$

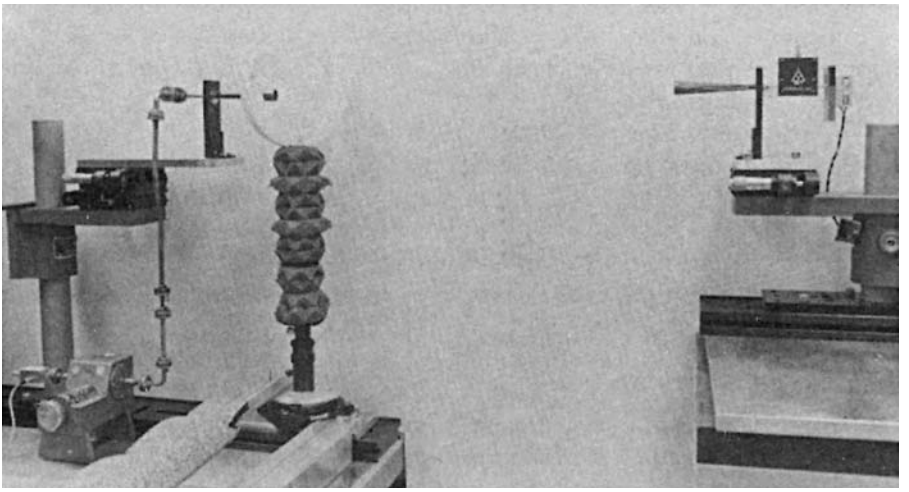


Figure 4.1 Photograph of experimental setup. The dielectric sample is mounted in the teflon holder. [From Shimabukuro et al. (1984).]

$$\beta_0 = \frac{2\pi}{\lambda_0}\cos\varphi,$$

$$\varepsilon_1 = \varepsilon_r\varepsilon_0\left(1 - \frac{i\sigma}{\omega\varepsilon_r\varepsilon_0}\right),$$

and

ε_0 = permittivity of free space
ε_r = relative permittivity
σ = conductivity
λ_0 = free-space wavelength
$\dfrac{\sigma}{\omega\varepsilon_r\varepsilon_0}$ = tan δ = loss tangent
d = thickness of the slab
r = reflection coefficient of a plane wave incident to a dielectric boundary
ω = free-space frequency
$i = \sqrt{-1}$

For more details on the various conditions of the experiment, see Shimbukuro et al. (1984).

We applied the bootstrap to the residuals using the nonlinear model

$$y_i = g_i(\mathbf{v}) + \varepsilon_i \qquad \text{for } i = 1, 2, \ldots, N$$

where y_i is the power transmission measurement at incident angle φ_i with $\varphi_i = i - 1$ degrees. The nonlinear function $g_i(u)$ is $|T|^2$ and $\mathbf{v}$ is a vector of two parameters, ε_r (relative permittivity) and tan δ (loss tangent). For simplicity the wavelength λ, the slab thickness d and the angle of incidence φ_i are all assumed to be known for each observation. The experimenters believe that measurement error in these variables would be relatively small and have little effect on the parameter estimates. Some checking of these assumptions was made.

For most of the materials, 51 observations were taken. We chose to do 20 bootstrap replications for each model. Results were given for eight materials and are shown in Table. 4.1.

The actual least-squares fit to the eight materials are shown in Figure 4.2. We notice that the fit is generally better at the higher-incidence angles. This suggests a violation of the assumption of independent and identically distributed residuals. There may be a bias at the low incidence angles indicative of either model inadequacy or poorer measurements.

Looking back on the experiment, there are several possible ways we might have improved the bootstrap procedure. Since bootstrapping residuals is more sensitive to the correctness of the model, it may have been better to bootstrap the vector.

Recent advances in bootstrapping in heteroscedastic models may also have helped. A rule of thumb for estimating standard errors is to take 100–200 bootstrap replications, whereas we only did 20 replications in this research.

Table 4.1 Estimates of Permittivities and Loss Tangents (f = 93.7888 GHz)

	Least-Squares Estimate		Bootstrap Estimates with Standard Error	
Material	ε_r	tan δ	ε_r	tan δ
Teflon	2.065	0.0002	2.065 ± 0.004	0.00021 ± 0.00003
Rexolite	2.556	0.0003	2.556 ± 0.005	0.00026 ± 0.00006
TPX	2.150	0.0010	2.149 ± 0.005	0.0009 ± 0.0001
Herasil (fused quartz)	3.510	0.0010	3.511 ± 0.005	0.0010 ± 0.0001
36D	2.485 (2.45)	0.0012 (<0.0007)	2.487 ± 0.008	0.0011 ± 0.0002
36DA	3.980 (3.7)	0.0012 (<0.0007)	3.980 ± 0.009	0.0014 ± 0.0001
36DK	5.685 (5.4)	0.0040 (<0.0008)	5.685 ± 0.009	0.0042 ± 0.0001
36DS	1.765 (1.9)	0.0042 (0.001)	1.766 ± 0.006	0.0041 ± 0.0001

Source: Shimabukuro et al. (1984).

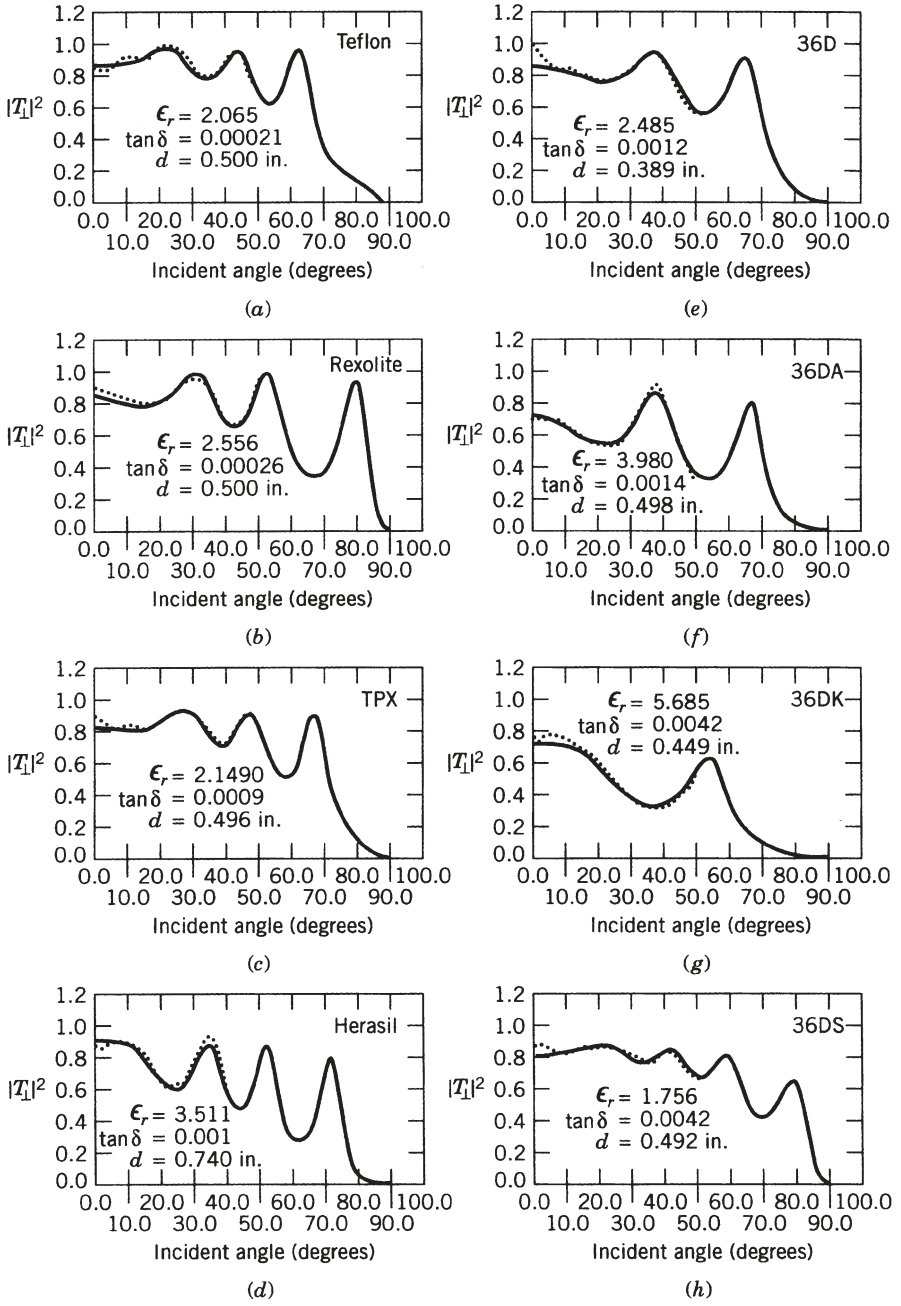


Figure 4.2 The measured power transmission for different dielectric samples is shown by the dotted lines. The line curves are the calculated $|T_\perp|^2$ using the best-fit estimates of ε_r and tan δ. [From Shimabukuro et al. (1984).]

From a data analytic point of view, it may have been helpful to delete the low-angle observations and see the effect on the fit. We might then have decided to fit the parameters and bootstrap only for angle greater than, say, 15 degrees.

By bootstrapping the residuals, the large residuals at the low angles would be added at the higher angles for some of the bootstrap samples. We believed that this would tend to increase the variability in the parameter estimates of the bootstrap sample and hence lead to an overestimate of their standard errors.

Since the estimated standard errors were judged to be good enough by the experimenters, we felt that our approach was adequate. The difficulty with the residual assumptions was recognized at the time.

4.3. NONPARAMETRIC MODELS

Given a vector $\boldsymbol{X}$, the regression function $E(y|\boldsymbol{X})$ is often a smooth function in $\boldsymbol{X}$. In Sections 4.1 and 4.2, we considered specific linear and nonlinear forms for the regression function. Nonparametric regression is an approach that allows more general smooth functions as possibilities for the regression function. The nonparametric regression model for an observed data set (y_i,x_i) for $1 \le i \le n$ is

$$y_i = g(x_i) + \varepsilon_i, \qquad 1 \le i \le n,$$

where $g(\boldsymbol{x}) = \boldsymbol{E}(y|\mathbf{x})$ is the function we wish to estimate. We assume that the ε_i are independent and identically distributed with mean zero and variance σ^2.

In the regression model, $\boldsymbol{x}$ is assumed to be given as in a designed experiment. One approach to the estimation of the function g is kernel smoothing [see Hardle (1990a,b) or Hall (1992a, pp. (257–269)]. The bootstrap is used to help determine the degree of smoothing (i.e., determine the tradeoff between variance and bias analogous to its use in nonparametric density estimation).

Cox's proportional hazards model is a standard regression method for dealing with censored data [see Cox (1972)]. The hazard function $h(t|\boldsymbol{x})$ is the derivative of the survival function $\boldsymbol{S}(t|\boldsymbol{x})$ = probability of surviving t or more time units given predictor variables $\boldsymbol{x}$. In Cox's model $h(t|\boldsymbol{x}) = h_0(t)e^{(\beta x)}$, where $h_0(t)$ is an arbitrary unspecified function assumed to depend solely on t.

Through the use of the "partial likelihood" function, the regression parameters β can be estimated independently of the function $h_0(t)$. Because of the form of $h(t|\boldsymbol{x})$, the method is sometimes referred to as semi parametric.

Efron and Tibshirani (1986) apply the bootstrap to leukemia data for mice in order to assess the effectiveness of a treatment. See their article for more details.

Without going into the details, we mention projection pursuit regression and alternating conditional expectation (ACE) as two other "nonparametric" regression techniques which have been studied recently. Efron and Tibshirani (1986) provide examples of applications of both methods and show how the bootstrap can be applied when using these techniques.

The interested reader can consult Friedman and Stuetzle (1981) for the original source on project pursuit. The original work describing ACE (or alternating conditional expectation) is Breiman and Friedman (1985).

Briefly, projection pursuit searches for linear combinations of the predictor variables and takes smooth functions of those linear combinations to form the prediction equation. ACE generalizes the Box–Cox regression model by transforming the response variable with an unspecified smooth function as opposed to a simple power transformation.

4.4. HISTORICAL NOTES

Although regression analysis is one of the most widely used statistical techniques, application of the bootstrap to regression problems has only appeared fairly recently. The many fine books on regression analysis including Draper and Smith (1981) for linear regression and Gallant (1987) and Bates and Watts (1988) do not mention or pay much attention to bootstrap methods. A recent exception is Sen and Srivastava (1990).

Draper and Smith (1998) also incorporate a discussion of the bootstrap. Early discussion of the two methods of bootstrapping in the nonlinear regression model with homoscedastic errors can be found in Efron (1982a). Carroll, Ruppert, and Stefanski (1995) deal with the bootstrap applied to the nonlinear calibration problem (measurement error models and other nonlinear regression problems, pp. 273–279, Appendix A.6).

Efron and Tibshirani (1986) provide a variety of interesting applications and some insightful discussion of bootstrap applications in regression problems. They go on to discuss nonparametric regression applications including projection pursuit regression and methods for deciding on transformations for the response variable such as the alternating conditional expectation method (ACE) of Breiman and Friedman (1985). Texts devoted to nonparametric regression and smoothing methods include Hardle (1990a,b), Hart (1997), and Simonoff (1996). Belsley, Kuh, and Welsch (1980) cover multicollinearity and related regression diagnostics.

Bootstrapping the residuals is an approach that also can be applied to time series models. We shall discuss time series applications in the next chapter. An example of a time series application to the famous Wolfer sunspot numbers is given in Efron and Tibshirani (1986, p. 65).

Shimabukuro et al. (1984) was an early example of a practical application of a nonlinear regression problem. The first major study of the bootstrap as applied to the problem of estimating the standard errors of the regression

coefficients by constrained least squares with an unknown, but estimated, residual covariance matrix can be found in Freedman and Peters (1984a). Similar analyses for econometric models can be found in Freedman and Peters (1984b).

Peters and Freedman (1984b) also deals with issues related to bootstrapping in regression problems. Their study is very interesting because it shows that the conventional asymptotic formulas that are correct for very large samples do not work well in small-to-moderate sample size problems. They show that these standard errors can be too small by a factor of nearly three! On the other hand the bootstrap method gives accurate answers. The motivating example is an econometric equation for the energy demand by industry.

In Freedman and Peters (1984b) the bootstrap is applied to a more complex econometric model. Here the authors show that the three-stage least-squares estimates and the conventional estimated standard errors of the coefficients are good. However, conventional prediction intervals based on the model are too small due to forecast bias and underestimation of the forecast variance.

The bootstrap approach given by Freedman and Peters (1984b) seems to provide better prediction intervals in their example. The authors point out that there is unfortunately no good rule of thumb to apply to determine when the conventional formulas will work of when it may be necessary to resort to the bootstrap. They suggest that the development of such a rule of thumb could be a result of additional research. Even the bootstrap procedure has problems in this context.

Theoretical work on the use of bootstrap in regression is given in Freedman (1981), Bickel and Freedman (1983), Weber (1984), Wu (1986), and Shao (1988a,b). Another application to an econometric model appears in Daggett and Freedman (1985).

Theoretical work related to robust regression is given in Shorack (1982). Rousseeuw (1984) applies the bootstrap to the least median of squares algorithm. Efron (1992a) discusses the application of bootstrap to estimating percentiles of a regression function.

Jeong and Maddala (1993) review various resampling tests for econometric models. Hall (1989c) shows that the bootstrap applied to regression problems can lead to confidence interval estimates that are unusually accurate.

Various recent regression applications include Breiman (1992) for model selection related to x-fixed prediction, Brownstone (1992) regarding admissibility of linear model selection techniques, Bollen and Stine (1993) regarding fitting of structural equation models, and Cao-Abad (1991) regarding rates of convergence for a bootstrap variation called the "wild" bootstrap. The wild bootstrap is useful in nonparametric regression [see also Mammen (1993), who applies the wild bootstrap in linear models], DeAngelis, Hall, and Young (1993a) related to L^1 regression, Lahiri (1994c) for M-estimation in multiple linear regression problems, Dikta (1990) for nearest-neighbor regression, and Green, Hahn, and Rocke (1987) for an economic application to the estimation of elasticities.

Wu (1986) gives a detailed theoretical treatment of jackknife methods applied to regression problems. He deals mainly with the problem of heteroscedastic errors. He is openly critical of the blind application of bootstrap methods and illustrates that certain bootstrap approaches will give incorrect results when applied to data for which heterosecedastic models are appropriate. A number of the discussants including Beran, Efron, Freedman, and Tibshirani defend the appropriate use of the "right" bootstrap in this context. The issue is a complex one which even today is not completely settled.

It is fair to say that Jeff Wu's criticism of the bootstrap in regression problems was a reaction to the "euphoria" expressed for the bootstrap in some of the earlier works such as Efron and Gong (1983, Section 1) or Diaconis and Efron (1983).

Although enthusiasm for the bootstrap approach is justified, some statements could leave naive users of statistical methods with the idea that it is easy to just apply the bootstrap to any problem they might have. I think that every bootstrap researcher would agree that careful analysis of the problem is a necessary step in any applied problem and that if bootstrap methods are appropriate, one must be careful to choose the "right" bootstrap method from the many possible bootstraps.

Stine (1985) deals with bootstrapping for prediction intervals, and Bai and Olshen as discussants to the paper by Hall (1988b) provide some elementary asymptotic theory for prediction intervals. Olshen, Biden, Wyatt, and Sutherland (1989) provide a very interesting application to gait analysis.

A theoretical treatment of nonparametric kernel methods in regression problems is given in Hall (1992a). His development is based on asymptotic expansions (i.e., Edgeworth expansions). Other key articles related to bootstrap applications to nonparametric regression include Hardle and Bowman (1988) and Hardle and Marron (1991).

The reader may first want to consult Silverman (1986) for a treatment of kernel density methods and some applications of the bootstrap in density estimation. Devorye and Gyorfi (1985) also deals with kernel density methods as does Hand (1982), and for multivariate densities see Scott (1992). Hardle (1990a) provides an account of nonparametric regression techniques.

Hayes, Perl, and Efron (1989) have extended bootstrap methods to the case of several unrelated samples with application to estimating contrasts in particle physics problems. Hastie and Tibshirani (1990) treat a general class of models called generalized additive models. These include both the linear and the generalized linear models as special cases. It can be viewed as a form of curve fitting but is not quite as general as nonparametric regression.

Bailer and Oris (1994) provide regression examples for toxicity testing and compare bootstrap methods with likelihood and Poisson regression models (a particular class of generalized linear models). One of their examples appears in Davison and Hinkley (1997, practical number 6, pp. 383–384).

CHAPTER 5

Forecasting and Time Series Analysis

5.1. METHODS OF FORECASTING

One of the most common problems in the "real world" is forecasting. We try to forecast tomorrow's weather or when the next big earthquake will hit. When historical data are available and models can be developed which fit the historical data well, we may be able to produce accurate forecasts. For certain problems (e.g., earthquake predictions or the Dow Jones Industrial Average) the lack of a good statistical model makes forecasting problematic (i.e., no better than crystal ball gazing).

Among the most commonly used forecasting techniques are exponential smoothing and autoregressive integrated moving average (ARIMA) modeling. The ARIMA models are often referred to as the Box–Jenkins models after George Box and Gwilym Jenkins, who popularized the approach in Box and Jenkins (1970, 1976). The autoregressive models, which are a subset of the ARIMA models, actually go back to Yule (1927).

Exponential smoothing is an approach that provides forecasts future values using exponentially decreasing weights on the past values. The weights are determined by smoothing constants that are estimated from the data. The simplest form—single exponential smoothing—is a special case of the ARIMA models namely the IMA (1, 1) model. The smoothing constant in the model can be determined from the moving average parameter of the IMA (1, 1) model. The smoothing constant can be determined from the moving average parameter of the IMA (1, 1) model.

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

5.2. TIME SERIES MODELS

ARIMA models are attractive because they provide good empirical approximations to a large class of time series. There is a body of statistical theory showing that "most" stationary stochastic processes can be well-approximated by high-order autoregressive processes.

The term *stationary stochastic process* generally means strictly stationary. A stochastic process is said to be strictly stationary if the joint probability distribution of k consecutive observations does not depend on the time parameter t for all choices of $k = 1, 2, 3, 4, 5, \ldots, \infty$. Informally, this means that if we are looking at the first k observations in a time series, the statistical properties of that set of observations wouldn't change if we took any other set of k consecutive observations in the time series.

A weaker form of stationarity is second-order (or weak) stationarity. Second-order stationarity requires only that the second-order moments exist and that the first- and second-order moments, the mean function and the autocorrelation function, respectively, do not depend on time (i.e., they are constant over time).

Strict stationarity implies weak stationarity, but there are weakly stationary, processes that are not strictly stationary. For Gaussian processes, second-order (weakly) stationary processes are strictly stationary because they have the property that the joint distribution, for any choice of k consecutive observations, depends only on the first and second moments of their joint distribution.

Box and Jenkins used the mixed autoregressive moving average model to provide a parsimonious representation for these high-order autoregressive processes (i.e., by including just a few moving average terms an equivalent model is found with only a small number of parameters to estimate). To generalize this further to handle trends and seasonal variations (i.e., nonstationarity), Box and Jenkins (1976) include differencing and seasonal differences of the series. Using mathematical operator notation, let

$$W_t = \Delta^d Y_t,$$

where Y_t is the original observation at time t and the operation Δ^d applies the difference operation Δd times where Δ is defined by $\Delta y_t = y_t - y_{t-1}$.

So,

$$\Delta^2 y_t = \Delta(y_t - y_{t-1}) = \Delta y_t - \Delta y_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}.$$

In general,

$$\Delta^d y_t = \Delta^{d-1}(\Delta y_t) = \Delta^{d-1}(y_t - y_{t-1}) = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}.$$

After differencing the times series, W_t is a stationary ARMA (p, q) model given by the equation

$$W_t = b_1 W_{t-1} + b_2 W_{t-2} + \cdots + b_p W_{t-p} + e_r + a_0 e_{t-2} + \cdots + a_q e_{t-q},$$

where e_t, $e_{t-1}, \ldots, e_{t-q}$ are the assumed random innovations and W_{t-1}, $W_{t-2}, \ldots, W_{t-p}$ are past values of the dth difference of the Y_t series.

These ARIMA models handle polynomial trends in the time series. Additional seasonal components can be handled by seasonal differences [see Box and Jenkins (1976) for details].

Although the Box–Jenkins models cover a large class of time series and provide very useful forecasts and prediction intervals, they have drawbacks for some cases. The models are linear and the least-squares or maximum likelihood parameter estimates are good only if the innovation series e_t is nearly Gaussian.

If the innovation series e_t has heavy tails or there are a few spurious observations in the data, the estimates can be distorted and the prediction intervals are not valid. In fact, the Box–Jenkins methodology for choosing the order of the model (i.e., deciding on the values for p, d, and q) will not work if outliers are present. This is because estimates for the autocorrelation and partial autocorrelation functions are very sensitive to outliers [see, for example, Chernick, Downing, and Pike (1982) or Martin (1980)].

One approach to overcoming the difficulty is to detect and remove the outliers and then fit the Box–Jenkins model with some missing observations. Another approach is to use robust estimation procedures for parameters [see Rousseeuw and Leroy (1987)].

In the 1980s there were also a number of interesting theoretical developments in bilinear and other nonlinear time series models which may help to extend the applicability of statistical time series modeling [see Tong (1983, 1990)].

Even if an ARIMA model is appropriate and the innovations e_t are uncorrelated but not Gaussian, it may be appropriate to bootstrap the residuals to obtain appropriate standard errors for the model parameters and the predictions. Bootstrap prediction intervals may also be appropriate.

The approach is the same as we have discussed in Chapter 4, which covers regression analysis. The confidence interval methods of Chapter 3 may be appropriate for the prediction intervals. We shall discuss this further in the next section.

5.3. WHEN DOES BOOTSTRAPPING HELP WITH PREDICTION INTERVALS?

Some results are available on the practical application of the bootstrap to time series models. These results apply to stationary autoregressive (AR)

processes, a subset of the stationary autoregressive-moving average (ARMA) models discussed in the previous section.

To illustrate how the bootstrap can be applied to an autoregressive model, we will illustrate the approach with the simple first-order autoregressive process. This model is sufficient to illustrate the key points. For the first-order autoregression (AR (1) model) the model is given by

$$y_t = b_1 y_{t-1} + e_t,$$

where y_t is the observation at time t (possible centered to have zero mean) and e_t are the innovations.

If the average of the observed series is not zero, a sample estimate of the mean is subtracted from each observation in order to center the data. In practice, if the original series appears to be nonstationary, differencing methods or other forms of trend removal would be applied first.

For Gaussian processes, least-squares or maximum likelihood estimates for b_1 are computed along with standard errors for the estimates. If y_{t_m} is the last observation, then a one-step-ahead prediction is obtained at $t_m + 1$ by using $\hat{b}_1 y_{t_m}$ as the prediction, where $\hat{b}_1$ is the estimate of b_1. Statistical software packages (e.g., SAS/ETS, BMDP, and IMSL) provide such estimates of parameters and also produce forecast intervals.

These procedures work well when the e_t have approximately a Gaussian distribution with mean zero. Stine (1987) provides forecasts and prediction intervals with the classical Gaussian model but using a bootstrap approach. He shows that although the bootstrap is not as efficient as the classical estimate when the Gaussian approximation is valid, it provides much better prediction intervals for non-Gaussian cases.

In order to apply the bootstrap to the AR (1) model, we need to generate a bootstrap sample. First we need an estimate $\hat{b}_1$. We may take the Gaussian maximum likelihood estimate generated by a software program such as PROC ARIMA from SAS. We then generate the estimated residuals, namely,

$$\hat{e}_i = y_t - \hat{b} y_{t-1} \qquad \text{for } t = 2, 3, \ldots, t_m.$$

Note that we cannot compute a residual $\hat{e}_1$ since y_0 is not available to us. A bootstrap sample $y_1^*, y_2^*, \ldots, y_{t_m}^*$ is then generated by bootstrapping the residuals. We simply generate $e_2^*, e_3^*, \ldots, e_{t_m}^*$ by sampling with replacement from $\hat{e}_2, \hat{e}_3, \ldots, \hat{e}_{t_m}$ and defining by recursion:

$$y_2^* = \hat{b}_1 y_1^* + e_2^*, \qquad y_3^* = \hat{b}_1 y_2^* + e_3^*, \ldots, y_{t_m}^* = \hat{b}_1 y_{t_m-1}^* + e_{t_m-1}^*.$$

Efron and Tibshirani (1986) take $y_1^* = y_1$ for each bootstrap sample. With autoregressive processes, since we have a first time point that we denote as $t = 1$, we need initial values. In the AR (1) example, we see that we need a single initial value to start the process. In this case we let $y_1^* = y_1$.

In general for the pth-order autoregression, we will need p initial values. Stine (1987) and Thombs and Schucany (1990) provide alternative methods for obtaining starting values for the bootstrap samples.

Now for each bootstrap sample, an estimate $\hat{b}_1^*$ is obtained by applying the estimation procedure to $y_1^*, y_2^*, \ldots, y_{t_m}^*$. Efron and Tibshirani illustrate this on the Wolfer sunspot data. They obtain the standard errors for $\hat{b}_1$ by this procedure. They then go on to fit an AR (2) (second-order autoregressive model) to the sunspot data and obtain bootstrap estimates of the standard errors for the two parameters in the AR (2) model. They did not go on to consider prediction intervals.

For the Gaussian case, the theory has been developed to obtain the minimum mean-square error predictions based on "known" autoregressive parameters. Formulas for the predictions and their mean-square errors can be found in Box and Jenkins (1976) or Fuller (1976). Stine (1987) shows that when the autoregressive parameter b_1 is replaced by the estimate $\hat{b}_1$ in the forecasting equations, the prediction mean-square error increases.

Stine (1987) provides a Taylor series expansion to estimate the mean-square error of the prediction that works well for Gaussian data. The bootstrap estimates of mean-square error are biased, but his bootstrap approach does provide good prediction intervals. We shall describe this approach, which we recommend when the residuals do not fit well to the Gaussian model.

Stine (1987) assumes that the innovations have a continuous and strictly increasing distribution with finite moments. He also assumes that the distribution is symmetric about zero. The key difference between Stine's approach and that of Efron and Tibshirani is the introduction of the symmetric error distribution. Instead of sampling with replacement from the empirical distribution for the estimated residuals (the method of Efron and Tibshirani previously described), Stine does the following:

Let

$$\begin{aligned} F_T(x) &= \frac{1}{2} + (L(x)/[2(T-p)]), \quad x \geq 0, t = p+1, \ldots, T \\ &= 1 - F_T(-x), \qquad\qquad\quad x < 0, \end{aligned}$$

where $L(x)$ = number of t such that $k|\hat{\varepsilon}_t| \leq x$, and

$$k = [(T-p)/(T-2p)]^{1/2}.$$

This choice of F_T produces bootstrap residuals that are symmetric about zero and have a variance that is the same as the original set of residuals.

A bootstrap approximation to the prediction error distribution is easily obtained given the bootstrap estimates of the autoregressive parameters and the bootstrap observations $y_1^*, y_2^*, \ldots, y_{t_m}^*$. The prediction formulas are used to obtain bootstrap prediction $\hat{y}_{t_m+f}^*$ for the time $t_m + f$, f time steps in the future. The variable $\hat{y}_{t_m+f}^* - \hat{y}_{t_m+f}$ provides the bootstrap sample estimate of prediction

error f steps ahead, where $\hat{y}^*_{t_m+f}$ is the original prediction based on the original estimates of the autoregressive parameters and the observations $y_1, y_2, \ldots, y_{t_m}$. Actually, Stine uses a more sophisticated approach based on the structure of the forecast equation [see Stine (1987) for details].

Another difference between Stine's approach and that of Efron and Tibshirani is that Efron and Tibshirani fix the first p values of the process in generating the bootstrap sample whereas Stine chooses a block of p consecutive observations at random to initiate the bootstrap sample.

In practice, we will know the last p observations when making future predictions. Autoregressive forecasts for 1, 2, . . . , f steps ahead depend only on the autoregressive parameters and the last p observations. Consequently, it makes sense to condition on the last p observations when generating the bootstrap predictions.

Thombs and Schucany (1990) use a time-reversal property for autoregressive processes to fix the last p observations and generate bootstrap samples for the earlier observations. They apply the backward representation (Box and Jenkins, 1976, pp. 197–200) to express values of the process at time t as a function of future values. This representation is based on generating the process backward in time, which is precisely what we want to do with the bootstrap samples. The correlation structure for the reversed process is the same as for the forward process.

For Gaussian processes, this means that the two series are distributionally equivalent. Weiss (1975) has shown that for linear processes (including autoregressions) the time-reversed version is distributional equivalent to the original only if the process is Gaussian.

Chernick, Daley, and Littlejohn (1988) provide an example of a first-order autoregression with exponential marginal distributions whose reversed version also has exponential marginals, is first-order Markov, and has a special structure. The process is not time-reversible (in the strict sense where reversibility means distribution equivalence of the two stochastic processes, original and time-reversed) as can be seen by looking at sample paths.

Thombs and Schucany (1990) also present simulation results that show that their method has promise. They did not use the symmetrized distribution for the residuals. In small samples, they concede that some refinements such as the bias-corrected percentile method might be helpful.

Unfortunately, we cannot recommend a particular bootstrap procedure as a "best" approach to bootstrapping time series even for generating prediction intervals for autoregressive time series. The method of Stine (1987) is recommended for use when the distributions are non-Gaussian. For nearly Gaussian time series, the standard methods available in most statistical time series programs are more efficient.

These methods are called model-based and the results do not work well when the form of the model is misspecified. Künsch (1989) was the first to develop the block bootstrap method in the context of stationary time series. It turns out to be a general approach that can be applied in many dependent

data situations including spatial data, *M*-dependent data, and time series. Lahiri has developed a theory for bootstrapping dependent data predominantly for classes of block bootstrap methods including (1) moving block bootstrap, (2) nonoverlapping block bootstrap, and (3) generalized block bootstrap that includes the circular block bootstrap and the stationary block bootstrap. This work is well-summarized along with other bootstrap methods for dependent data in the text by Lahiri (2003a). Block-based versus model-based bootstrap methods are considered in the next section. Alternative approaches to time series problems are described in Sections 5.4 and 5.5, with block resampling methods contrasted to model-based methods in Section 5.4.

5.4. MODEL-BASED VERSUS BLOCK RESAMPLING

The methods described thus far all fall under the category of model-based resampling methods, because the residuals are generated and resampled based on a time series model [i.e., the AR(1) model in the earlier illustration]. Refinements to the above approach are described in Davison and Hinkley (1997, pp. 389–391).

There they center the residuals by subtracting the average of the residuals. They then use a prescription just as we have described above. However, they point out that the generated series is not stationary. This is due to the initial values. This could be remedied by starting the series in equilibrium or more practically by allowing a "burn-in" period of k observations that are discarded. We choose k so that the series has "reached" stationarity.

To use the model-based approach, we need to know the parameters and the structure of the model, and this is not always easy to discern from the data. If we choose an incorrect structure, the resampled series will have a different structure (which we incorrectly thrust upon it) from the original data and hence will have different statistical properties. So if we know that we have a stationary series but we don't know the structure, analogous to the nonparametric alternative to the distributional assumptions for the observations, we would like a bootstrap resampling structure that doesn't depend on this unknown structure. But what is the time series analog to nonparametric models?

Bose (1988) showed that if an autoregressive process is a "correct" model (or for practical use at least approximately correct) there is an advantage to using the model-based resampling approach, namely, good higher-order asymptotic properties for a wide variety of statistics that can be derived from the model. On the other hand, we could pay a heavy price, in that the estimates could be biased and/or grossly inaccurate if the model structure is wrong. This is very much like the tradeoff we have between parametric and nonparametric inference where the model is the assumed parametric family of distributions for the observations.

A remedy, the block bootstrap, which was first introduced by Carlstein (1986), was further developed by Künsch (1989) and is a method that resamples the time series, in blocks (possibly overlapping blocks). For uncorrelated exchangeable sequences, the original nonparametric bootstrap that resamples the individual observations is appropriate. For stationary time series, successive observations are correlated but observations separated by a large time gap are nearly uncorrelated. This can be seen by the exponentially declining autocorrelation function for a stationary AR (1) model.

A key idea in the development and success of block resampling is that, for stationary series, individual blocks of observations that are separated far enough in time will be approximately uncorrelated and can be treated as exchangeable. So suppose the time series has length $n = bl$. We can generate b nonoverlapping blocks each of length l.

The key idea that underlies this approach is that if the blocks are sufficiently long, each block preserves, in the resampled series, the dependence present in the original data sequence. The resampling or bootstrap scheme here is to resample with replacement from the set of b blocks.

There are several variants on this idea. One is to allow the blocks to overlap. This was one of Künsch's proposals, and it allows for more blocks than if they are required not to overlap.

Suppose we take the first block to be (y_1, y_2, y_3, y_4), the second to be (y_2, y_3, y_4, y_5), the third to be (y_3, y_4, y_5, y_6), and so on. The effect of this approach is that the first $l - 1$ observations from the original series appear in fewer blocks than the rest.

Note that observation y_1 appears in only one block, y_2 appears in only two blocks, and so on. This effect can be overcome by wrapping the data around in a circle (i.e., the last observation in the series is followed again by the first, etc.)

At the time of the writing of the first edition the block bootstrap approach was the subject of much additional research. Many theoretical results and applications have occurred from 1999 to the present (2007).

Professor Lahiri, from Iowa State University, has been one of the prime contributors and has nicely summarized the theoretical properties and (through examples) the applications of the various types of block bootstrap methods for time series and other models of dependent data (including spatial data) in his text (Lahiri, 2003a). I will not cover these topics in depth but rather refer the reader to the literature and the chapters in the Lahiri text as they are discussed.

The various block bootstraps discussed in Lahiri (2003a, Chapter 2) are (1) the moving block bootstrap (MBB), (2) nonoverlapping block bootstrap (NBB), (3) circular block bootstrap (CBB), and (4) the stationary block bootstrap (SBB). I will give formal definitions and discuss these methods in detail later in this section.

Lahiri (2003a) also compares various block methods based on both theory and empirical simulation results (Chapter 5), covers methods for selecting the

block size for the moving block bootstrap (Chapter 7), pointing out how it can be generalized to other block methods, and covers model-based methods (Chapter 8) including the ones discussed in this chapter and more, frequency domain methods (Chapter 9) such as the ones we will discuss in Section 5.5, long-range-dependent models (Chapter 10), heavy-tailed distributions and the estimation of extreme values (Chapter 11), and spatial data (Chapter 12). In Chapter 8 we will cover some of the results for spatial data and in Chapter 9 we will cover situations where the naïve bootstrap fails, which includes the estimation of extreme values. Lahiri shows that for dependent data, the moving block bootstrap also fails if the resample size m is the same as the original sample size n. But as we will see for the independent case in Chapter 9 of this text, an m-out-of-n bootstrap remedies the situation. Lahiri derives the same result for MBB.

Some of the drawbacks of block methods in general are as follows: (1) Resampled blocks do not quite mimic the behavior of the time series, and (2) they have a tendency to weaken the dependency in the series.

Two methods, postblackening and resampling blocks of blocks, both help to remedy these problems. The interested reader should consult Davison and Hinkley (1997, pp. 397–398) for some discussion of these methods.

Another simple way to overcome this difficulty is what is called the stationary block bootstrap, SBB as referred to by Lahiri (2003a) and described in Section 2.7.2 of Lahiri (2003a) with statistical properties for the sample mean given in Section 3.3 of Lahiri (2003a). The stationary block bootstrap is a block bootstrap scheme that instead of having fixed length blocks has a random block length size. The distribution for block length is given using the random length L, where

$$\Pr(L = j) = (1 - p)^{j-l} p, \qquad \text{for } j = 1, 2, 3, \ldots, \infty.$$

This length distribution is the geometric distribution with parameter p. The mean block length for L is $\lambda = p^{-1}$. We may choose λ as one might choose the length of a fixed block length. Since $\lambda = 1/p$, determining λ also determines p. The stationary block bootstrap was first described by Politis and Romano (1994a).

It appears that the block resampling method has desirable properties of robustness to model specification in that it applies to a broad class of stationary series. Other variations and some theory related to block resampling can be found in Davison and Hinkley (1997, pp. 401–403) for choice of block length and pp. 405–408 for the underlying theory. Hall (1998) provides an overview of the subject. A very detailed and up-to-date coverage of block resampling can be found in the text Lahiri (2003a) and in the summary article Lahiri (2006) in the book "Frontiers in Statistics" Fan and Koul (2006).

Davison and Hinkley (1997) illustrate the application of block resampling using data on the river heights over time for the Rio Negro. A concern of the study was that there is a trend for heights of the river near Manasas to increase

over time due to deforestation. A test for trend was applied, and there is some evidence that a trend may be present but the statistical test was inconclusive. The trend test was based on a test statistic that was a linear combination of the observations, namely,

$$T = \sum a_i Y_i,$$

where, for $i = 1, 2, 3, \ldots, n$, Y_i is the sequence for river levels at Manasas and

$$a_i = (-1)[1 - ((i-1)/(n+1))]^{1/2} - i[1 - i/(n+1)]^{1/2} \qquad \text{for } i = 1, 2, 3, \ldots, n.$$

The test based on this statistic is optimal for detecting a monotonic trend when the observations are independent and identically distributed (i.e., IID under the null hypothesis). However, the time series data show clear autocorrelation at time lags *i*. A smoothed version of the Rio Negro river heights (i.e., a centered ten-year moving average) is shown in Figure 5.1 taken from Davison and Hinkley (1997).

The test statistic T above is still used, and its value in the example turns out to be 7.908. But is this statistically significantly large based on the null hypothesis? Instead of using the distribution of the test statistic under the null hypothesis, Davison and Hinkley choose to estimate its null distribution using block resampling. This is a more realistic approach for the Rio Negro data.

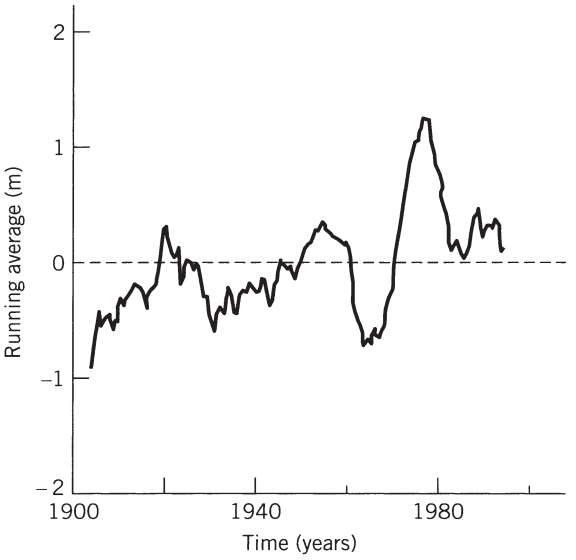


Figure 5.1 Ten-year running average of the Manasas data. [From Davison and Hinkley (1997, Figure. 8.9, p. 403), with permission from Cambridge University Press.]

They compare the stationary bootstrap to a fixed block length method. The purpose is to use the bootstrap to estimate the variance of T under the null hypothesis that the series is stationary but uncorrelated (as opposed to an IID null hypothesis). The asymptotic normality of T is used to do the statistical inference.

Many estimates were obtained using these two methods because various block sizes were used. For the fixed block length method, various fixed block sizes were chosen; for the stationary bootstrap, several average block lengths were specified. The bottom line is that the variance of T is about 25 based on the first 120 time points, but the lowest "reasonable" estimate for the variance of T based on the entire series is approximately 45! This gives us a p-value of 0.12 for the test statistic, indicating a lack of strong evidence for a trend.

When considering autoregressive processes, there are three cases to consider that involve the roots of the characteristic polynomial associated with the time series. See Box and Jenkins (1976) for details about the characteristic polynomial and the relationship of its roots to stationarity. The roots of the characteristic polynomial are found in the complex plane. If all the roots fall inside the unit circle, the time series is stationary. When one or more of the roots lies on the boundary of the unit circle, the time series is nonstationary and called unstable. If all the roots of the characteristic polynomial lie outside the unit circle, the time series is nonstationary and called explosive. In the first case the model-based method that Lahiri calls the autoregressive bootstrap (ARB) can be used. In the case of unstable processes the ARB bootstrap is not consistent but can be made consistent by an m-out-of-n modification that is one of the two. In the case of explosive processes, another remedy is required. The details are given in Chapter 8 of Lahiri (2003a) and the remedies are also covered in Chapter 9, where we cover remedies when the ordinary bootstrap methods fail.

5.5. EXPLOSIVE AUTOREGRESSIVE PROCESSES

An explosive autoregressive process is simply an autoregressive time series whose characteristic polynomial has all its roots outside the unit circle. As such, it is nonstationary process with unusual properties.

Datta (1995) showed that the normalized least-squares estimator of the autoregressive parameters in the explosive case converges to a nonnormal limiting distribution that is dependent on the initial p-observations. As a result, in the explosive case, any bootstrap method needs to use a consistent estimate of joint distribution of the first p-observations. Or alternatively, one can consider the distribution of the parameter by conditioning on the first p-observations. This is how Lahiri (2003a) constructs a consistent ARB estimate. In the explosive case the innovation series may not have a finite expectation; so although in the stationary case the innovations are centered, they cannot be in the explosive case.

The bootstrap observations are generated by the following bootstrap recursion relationship:

$$X_i^* = \hat{\beta}_{1n} X_{i-1}^* + \cdots + \hat{\beta}_{pn} X_{i-p}^* + \varepsilon_i^*, \qquad i \geq p+1.$$

This is well-defined when because of the conditioning argument we set $(X_1^*, \ldots, X_p^*)' \equiv (X_1, \ldots, X_p)'$. The bootstrap error variables ε_i^* are generated at random with replacement from the residuals, $\{\varepsilon_i \equiv X_i - \sum_{j=1}^{p} \beta_{jn} X_{i-j}$: $p+1 \leq i \leq n\}$. Datta has proven [Theorem 3.1 of Datta (1995)] that this ARB is consistent. This result may seem surprising since in the unstable case a similar ARB is not consistent and requires an m-out-of-n bootstrap to be consistent.

5.6. BOOTSTRAPPING STATIONARY ARMA PROCESSES

The stationary ARMA process was first popularized by Box and Jenkins (1970) as a representation that is parsimonious in terms of parameters. The process could also be represented as an infinite moving average process or possibly even an infinite autoregressive process. In practice, since the process is stationary, the series could be approximated by a finite AR process or a finite moving average process. But in either case the number of parameters required in the truncated process is much more than the few AR and MA parameters that appear in the ARMA representation.

Now let $\{X_i\}$, $i \in Z$, be a stationary ARMA (p, q) process satisfying the equation

$$X_i = \sum_{j=1}^{p} \beta_j X_{i-j} + \sum_{j=1}^{q} \alpha_j \varepsilon_{i-j} + \varepsilon_i, \qquad i \in Z,$$

where p and q are integers greater than or equal to 1. The formal description of this model-based bootstrap is involved but can be found in Lahiri (2003a, pp. 214–217). He invokes the standard stationarity and invertibility conditions that Box and Jenkins (1970) generally assume for an ARMA process. Given these conditions, the ARMA process admits both an infinite moving average and an infinite autoregressive representation. The resulting bootstrap is called ARMAB by Lahiri.

5.7. FREQUENCY-BASED APPROACHES

As we have mentioned before, second-order stationary Gaussian processes are strictly stationary as well and are characterized by their mean value function and their autocovariance (or autocorrelation) function. The Fourier transformation of the autocorrelation function is a function of frequency called the spectral density function.

Since a mean zero stationary Gaussian process is characterized by its autocorrelation function and the Fourier transform of the autocorrelation function is invertible, the spectral density function also characterizes the process. This helps explain the importance of the autocorrelation function and the spectral density function in the theory of stationary time series (especially for stationary Gaussian time series). Time series methods based on knowledge or estimates of the autocorrelation function are called time domain methods, and time series methods based on the spectral density function are called frequency domain methods. Brillinger (1981) gives a nice theoretical account of the frequency domain approach to time series.

The periodogram, the sample analog to the spectral density function, and smoothed versions of the periodogram that are estimates of the spectral density function have many interesting and useful properties, which are covered in detail in Brillinger (1981). The Fourier transform of the time series data itself is a complex function called the empirical Fourier transform.

From the theory of stationary processes, it is known that if the process has a well-defined spectral density function and can be represented by an infinite moving average process (representation), then as the series length $n \to \infty$ the real and imaginary parts of this empirical Fourier transform at the Fourier frequencies $\omega_\kappa = 2\pi\kappa/n$ are approximately independent and normally distributed with mean zero and variance $ng(\omega_\kappa)/2$, where $g(\omega_\kappa)$ is the true spectral density function at ω_κ.

This asymptotic result is important and practically useful. The empirical Fourier transform is easy to compute thanks to a technique known as the fast Fourier transform (FFT), and independent normal random variables are easier to deal with than nonnormal correlated variables.

So we use these ideas to construct a bootstrap. Instead of bootstrapping, the original series we can use a parametric bootstrap on the empirical Fourier transformed data. In the frequency domain we have basically an uncorrelated series of observations on the set of Fourier frequencies. The parametric bootstrap samples the indices of the Fourier frequencies with replacement, and then at each sampled frequency a bootstrap observation is generated from the estimated normal distribution. This generates a bootstrap version of the empirical Fourier transform, and then a bootstrap sample for the original series is obtained by inverting this Fourier transform. This idea has been exploited in what Davison and Hinkley (1997) call the phase scrambling algorithm. Although the concept is easy to understand, the actual algorithm is somewhat complicated. The interested reader can see more detail and examples in Davison and Hinkley (1997, pp. 408–409).

Davison and Hinkley (1997) then apply the phase scrambling algorithm to the Rio Negro data. This allows them to compare their previous time domain bootstrapping approach (SSB) with this frequency domain approach. For the null hypothesis, they again assume that the series is an AR (2) process and get an estimate of the variance of the trend estimator T. Using the frequency domain approach, they again determine T to be close to 51. So this result is very close to the result from the previous time domain SSB approach.

Now under the conditions described above, the periodogram has its values at the Fourier frequencies, and they are well-approximated as independent identically distributed exponential random variables. If one is interested only in confidence intervals for the spectral density at certain frequencies or to access variability of estimates that are based on the periodogram values, it is only necessary to resample the periodogram values and you don't have to bother with the empirical Fourier transform or the original time series. This method is called periodogram resampling, and details about the method and its applications to inference about the spectral density function are given by Davison and Hinkley (1997, pp. 412–414).

These frequency domain bootstraps are part of a general category of methods called transformation-based bootstraps where the bootstrapping all takes place on the transformed data and analysis can then be done in the time domain after taking the inverse transform. Lahiri (2003a) covers a number of these approaches on pages 40–41 of the text and uses the acronym TBB for transformation-based bootstrap. Lahiri provides a generalization of a method due originally to Hurvich and Zeger (1987) which is similar conceptually but still different from the method described above from the Davison and Hinkley (1997) text.

Hurvich and Zeger (1987) consider the discrete Fourier transform (DFT) of the data and bootstrap the transformed data rather the original series and then apply the IID nonparametric bootstrap to this transformed data. In this way, they also take advantage of the result in time series analysis that the Fourier transform of the series at distinct frequencies λ_i, where $-\pi < \lambda_i \leq \pi$, are approximately distributed as complex normal and are independent [see Brillinger (1981) or Brockwell and Davis (1991, Chapter 10) for more details].

In Lahiri (2003a), he generalized the approach of Hurvich and Zeger. His development now follows. We let $\theta = \theta(P)$ be the parameter of interest and P the probability measure that generates the observed series. Let T_n be an estimator of θ based on the observed series up to time n. The goal is to approximate the sampling distribution of a studentized statistic R_n that is used to draw inference about θ. The bootstrapping is done on R_n and will be used to get estimates of θ. See Lahiri (2003a, pp. 40–41) and Lahiri (2003a, Chapter 9) for further discussion of the Hurvich and Zeger approach along with more detail about the use of frequency domain bootstraps (FDBs).

5.8. THE SIEVE BOOTSTRAP

Another time domain approach to bootstrap from a stationary stochastic process is called the sieve bootstrap. We let P be the unknown joint probability distribution of the "infinite time series sequence" $\{X_1, X_2, X_3, \ldots, X_n, \ldots\}$. In the IID case we use the empirical distribution F_n or some other estimate of the marginal distribution F and the joint distribution for the first n observations is the product of the F_n s by independence. In this case, because the

observations in the time series are dependent, the joint distribution is not the product of the marginal distributions.

The idea of the sieve bootstrap is to choose a sequence of joint distributions $[\tilde{P}_n]_{n>0}$ called a sieve that approximates P. This sequence is such that for each n the probability measure $\tilde{P}_{n+1}$ is a finer approximation to P than the previous member of the sequence $\tilde{P}_n$. This sequence of measures converges to P as $n \to \infty$ in an appropriate sense.

For a large class of stationary processes, Bühlmann (1997) presents a sieve bootstrap method based on a sieve of increasing order, a p^{th}-order autoregressive process. Read Bühlmann (1997) for more details. We will give a brief description similar to the description in Lahiri (2003a). Another approach suggested in Bühlmann (2002a) is based on a variable-length Markov chain. When considering the choice of a sequence of approximating distributions for the sieve, there is a tradeoff between the accuracy of the approximating distribution and its range of validity. This tradeoff is discussed in Lahiri (2002b).

Now let us consider a stationary sequence $[X_n]$ with $n \in Z$, where Z is the set of positive integers with $EX_1 = \mu$ that admits a one-sided infinite moving average representation given by

$$X_i - \mu = \sum_{j=0}^{+\infty} \alpha_j \varepsilon_{i-j}, \qquad i \in Z$$

with $\sum_{j=1}^{+\infty} \beta_j^2 < \infty$. This representation indicates that for autoregressive processes of finite order: $p_n \to \infty$ as $n \to \infty$, but $n^{-1} p_n \to 0$ as $n \to \infty$. The autoregressive representation is given by

$$X_i - \mu = \sum_{j=1}^{p_n} \beta_j (X_{i-j} - \mu) + \varepsilon_i, \qquad i \in Z.$$

Using the autoregressive representation above, we fit the parameters β_j to an AR (p_n) model. The sieve is then based on the sequence of probability measures associated with the fitted AR (p_n) model. For more details see Lahiri (2003a, pp. 41–43). In his paper, Bühlmann (1997) establishes the consistency of this autoregressive sieve bootstrap.

5.9. HISTORICAL NOTES

The use of ARIMA and seasonal ARIMA models for forecasting and control problems was first popularized by Box and Jenkins (1970, 1976). This work was recently updated in Box, Jenkins, and Reinsel (1994). A classic theoretical text on time series analysis is Anderson (1971).

A popular common theoretical account of time series analysis is Brockwell and Davis (1991), which covers both time domain and frequency domain

analysis. Fuller (1976) is another excellent text at the high undergraduate or graduate school level that also covers both domains well. A couple of articles by Tong (1983, 1990) deal with nonlinear time series models.

Bloomfield (1976), Brillinger (1981), and Priestley (1981) are all time series texts that concentrate strictly on the frequency domain approach. Hamilton (1994) is another major text on time series. Braun and Kulperger (1997) did some work on the Fourier transform approach to bootstrapping.

The idea of bootstrapping residuals was described in Efron (1982a) in the context of regression. It is not clear who was the first to make the obvious extension of this to ARMA time series models. Findley (1986) was probably the first to point out some difficulties with the bootstrap approach particularly regarding the estimation of mean-square error.

Efron and Tibshirani (1986) showed how bootstrapping residuals provided improved standard error estimates for the autoregressive parameter estimates for the Wolfer sunspot data. Stine (1987) and Thombs and Schucany (1990) provide refinements to obtain better prediction intervals. Other empirical studies are Chatterjee (1986) and Holbert and Son (1986). McCullough (1994) provides an application of bootstrapping prediction intervals for AR(p) models.

Results for nonstationary autoregressions appear in Basawa, Mallik, McCormick, and Taylor (1989) and Basawa, Mallik, McCormick, Reeves, and Taylor (1991a,b). Theoretical developments are given in Bose (1988) and Künsch (1989).

Künsch (1989) is an attempt to develop a general theory for bootstrapping stationary time series. Bose (1988) also shows good asymptotic higher-order properties when applying model-based resampling to a wide class of statistics used with autoregressive processes.

Shao and Yu (1993) apply the bootstrap for the sample mean in a general class of time series, namely, stationary mixing processes. Hall and Jing (1996) apply resampling methods to general dependent data situations. Lahiri (2003a) is the new authoritative and up-to-date text to cover time series and other dependent data problems, including extremes in stationary processes and spatial data models.

Model-based resampling for time series was discussed by Freedman (1984), Freedman and Peters (1984a,b), Swanepoel and van Wyk (1986), and Efron and Tibshirani (1986). Li and Maddala (1996) provide a survey of related time domain literature on bootstrapping with emphasis on econometric applications.

Peters and Freedman (1985) deal with bootstrapping for the purpose of comparing competing forecasting equations. Tsay (1992) provides an applied account of parametric bootstrapping of time series.

Kabaila (1993a) discusses prediction in time series. Stoffer and Wall (1991) apply the bootstrap to state space models for time series. Chen, Davis, Brockwell, and Bai (1993) use model-based resampling to determine the appropriate order for an autoregressive model.

Good higher-order asymptotic properties for block resampling [similar to the work in Bose (1988)] have been demonstrated by Lahiri (1991) and Götze and Künsch (1996). Davison and Hall (1993) show that good asymptotic properties for the bootstrap generally depend crucially on the choice of a variance estimate. Lahiri (1992b) applies an Edgeworth correction in using the moving block bootstrap for both stationary and nonstationary time series models.

Block resampling was introduced by Carlstein (1986). The key breakthrough with the block resampling approach came later when Künsch (1989) provided many of the important theoretical developments on the block bootstrap idea and introduced the idea of overlapping blocks.

The stationary bootstrap was introduced by Politis and Romano (1994a). They also proposed the circular block bootstrap in an earlier work, Politis and Romano (1992a). Liu and Singh (1992b) obtain general results for moving block jackknife and bootstrap approaches to general types of weak dependence. Liu (1988) and Liu and Singh (1995) deal with bootstrap approaches to general data sets that are not IID. For the most recent developments in block bootstrap theory and methods see Lahiri (2003a) and Lahiri (2006).

Theoretical developments for general block resampling schemes followed the work of Künsch, in the articles Politis and Romano (1993a, 1994b), Bühlmann and Künsch (1995), and Lahiri (1995). Issues of block length are addressed by Hall, Horowitz, and Jing (1995). Lahiri (2003a, pp. 175–186) covers optimal block sizes for estimating bias, variance, and distribution quantiles. He covers much of the research from Hall, Horowitz, and Jing (1995).

Fan and Hung (1997) use balanced resampling (a variance reduction technique that is covered in Chapter 7) to bootstrap finite Markov chains. Liu and Tang (1996) use bootstrap method for control charting in both the independent and dependent situations.

Frequency domain resampling has been discussed by Franke and Hardle (1992) with an analogy to nonparametric regression. Janas (1993) and Dahlhaus and Janas (1996) extended these results. Politis, Romano, and Lai (1992) provide bootstrap confidence bands for spectra and cross-spectra (frequency domain analog, respectively, autocorrelation and cross-correlation functions in the time domain).

The sieve bootstrap was introduced for a class of stationary stochastic processes (that admit an infinite moving average representation) by Bühlmann (1997). It is also covered in Section 2.10 of Lahiri (2003a).

CHAPTER 6

Which Resampling Method Should You Use?

Throughout the first five chapters of this book, we have discussed the bootstrap and many variations in many different contexts, including point estimation, confidence intervals, hypothesis tests, regression problems, and time series predictions. In addition to considering whether or not to use the empirical distribution, a smoothed version of it, or a parametric version of it, we also considered improvements on bootstrap confidence interval estimates through the bias correction and acceleration constant adjustment to Efron's percentile method or by bootstrap iteration of any particular bootstrap confidence interval or by the use of the bootstrap percentile *t* method.

In some applications, related resampling techniques such as the jackknife, cross-validation, and the delta method have been considered. These three other resampling methods have been the traditional solutions to the problem of estimating the standard error of an estimate and were introduced before the bootstrap.

In the case of linear regression with heteroscedastic errors, Wu (1986) pointed out problems with the standard bootstrap approach and offered more effective jackknife estimators. Other authors have proposed other variants to the bootstrap, which they claim work just as well as the jackknife in the heteroscedastic case.

In the case of error rate estimation in discriminant analysis, Efron (1983) showed convincingly that the bootstrap and some variants (particularly the .632 estimator) are superior to cross-validation. Other simulation studies supported and extended these results.

With regard to bootstrap sampling, in Chapter 7 we shall illustrate certain variance reduction techniques that help to reduce the number of bootstrap

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

resamples (iterations) needed to adequately approximate the bootstrap estimate. Sometimes variants of the bootstrap are merely applications of a different variance reduction method.

With all these variants of the bootstrap and related resampling techniques, the practitioner may naturally wonder which of these various techniques should be applied to his or her particular problems. The answer may or may not be clear-cut, depending on the application. Also, because the research work on the bootstrap is still maturing, the jury is still out on some of these real-world problems.

Nevertheless, the purpose of this chapter is to sort out the various resampling techniques and describe them for the practitioner. We discuss their similarities and differences and, where possible, recommend the preferred techniques.

The title of the chapter is intended to be provocative. The chapter will not provide a complete answer to the question raised in the title, but, when possible, an answer is provided. Unfortunately, in many situations there is still no definitive answer.

6.1. RELATED METHODS

6.1.1. Jackknife

As pointed out in Section 2.1, the jackknife goes back to Quenouille (1949), whose goal it was to improve an estimate by correcting for bias. Later it was discovered that the jackknife is even more useful as a way to estimate variances or standard errors of estimators.

In general, we consider an estimate $\hat{\varphi}$ based on a sample $x_1, x_2, \ldots, x_n$ of observations that are independently drawn from a common distribution F. Suppose that $\hat{\varphi}$ can be represented as a functional of F_n, the empirical distribution [i.e., $\varphi = \varphi(F)$ and $\hat{\varphi} = \varphi(F_n)$]. Now we define $\hat{\varphi}_{(i)} = \varphi(F_{n(i)})$, where $F_{n(i)}$ places probability mass $1/(n-1)$ on the observations $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ and no mass on x_i. The jackknife estimate of variance is then defined as

$$\hat{\sigma}^2_{JACK} = \left(\frac{n}{n-1}\right)\sum_{i=1}^{n}[\hat{\varphi}_{i*} - \hat{\varphi}_*]^2,$$

where

$$\hat{\varphi}_* = \frac{1}{n}\sum_{i=1}^{n}\hat{\varphi}_{(i)}.$$

The jackknife estimate of the standard error of $\hat{\varphi}$ is just the square root of $\hat{\sigma}^2_{JACK}$.

Tukey, whose work came after that of Quenouille (1949), defined a quantity

$$\hat{\varphi}_i = \hat{\varphi} + (n-1)(\hat{\varphi} - \hat{\varphi}_{(i)}),$$

which he called the ith pseudo-value. The reason for this is that for general statistics, $\hat{\varphi}$ (the jackknife estimate of φ) is given by

$$\hat{\varphi} = \sum_{i-1}^{n} \hat{\varphi}_i / n$$

and

$$\hat{\sigma}^2_{JACK} = \sum_{i-1}^{n} (\hat{\varphi}_i - \hat{\varphi})^2 / [n(n-1)].$$

This is the standard estimate used for the variance of a sample mean (in this case the sample mean of the pseudo-values).

The jackknife has proven to be a very useful tool in the estimation of the variance of more complicated estimators such as robust estimators of location like trimmed and Winsorized means [see Efron (1982a, pp. 14–16) for details and discussion]. Simulation studies by Efron have generally shown the bootstrap estimate of standard deviation to be superior to the jackknife [see the trimmed means example in Efron (1982a, pp. 15–16 and Chapter 6), and for an adaptive trimmed mean see Efron (1982a, pp. 28–29)].

In Section 2.2.2 we provided a bootstrap estimate of the standard error for the sample median. The jackknife prescription provides an estimate that is not even consistent [see Efron (1982a, p. 16 and Chapter 6) for details)] in the case of the sample median.

All this empirical and theoretical evidence leads us to a recommendation to use the bootstrap instead of the jackknife when determining a standard error for an estimator. Also in Efron (1982a), Theorem 6.1, he shows that the jackknife estimate of a standard error is a bootstrap estimate with $\hat{\varphi}$ replaced by a linear approximation [up to a factor $\sqrt{n/(n-1)}$]. This result suggests that the jackknife estimate is an approximation to the bootstrap, and some researchers use this as another point to argue in favor of the bootstrap. Beran (1984a) determines jackknife approximations to bootstrap estimates exploiting some of the ideas posited by Efron.

6.1.2. Delta Method, Infinitesimal Jackknife, and Influence Functions

Many times we may be interested in the moments of an estimator (e.g., for variance estimation, the second moment). In such cases, it may be difficult to derive the exact moments. Nevertheless, the estimator may be represented as

a function of other estimators whose first moments are known. As an example, the correlation coefficient ρ between random variables X and Y is defined as

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}},$$

where $\mathrm{Cov}(X,Y)$ is the covariance between X and Y and, $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ are the respective variances of X and Y. The method described here, known as the delta method, is often used in such situations and particularly in simple cases where we want the variance of a transformed variable such as X^p or $\log(X)$.

To illustrate, assume $\varphi = f(\alpha)$, where φ and α are one-dimensional variables and f is differentiable with respect to α. The procedure to be described can be generalized to multidimensional φ and α with $\mathbf{f}$ a vector-valued function. Viewing α as a random variable with expected value α_0, we produce a first-order Taylor series expansion of φ about $f(\alpha_0)$. So

$$\varphi = f(\alpha) - f(\alpha_0) + (\alpha - \alpha_0) f'(\alpha_0) + \text{remainder terms}$$

and dropping the remainder terms we have

$$f(\alpha) - f(\alpha_0) \approx (\alpha - \alpha_0) f'(\alpha_0).$$

Upon squaring both sides of the above equation and taking expectations, we have

$$E((f(\alpha) - f(\alpha_0))^2 \approx E(\alpha - \alpha_0)^2 [f'(\alpha_0)]^2. \tag{6.1}$$

Now $E(\alpha - \alpha_0)^2$ is the variance of α and $f'(\alpha_0)$ is known. The left-hand side of the above equation is approximately the variance of $f(\alpha)$ or φ.

In the case where the variance of α is unknown, Efron (1982a, p. 43) suggests the nonparametric delta method where formulas like Eq. (6.1) are derived but applied to the empirical distribution for α [i.e., in the case of Eq. (6.1), the sample estimate of variance for α replaces $E(\alpha - \alpha_0)^2$].

Using geometrical ideas, Efron (1982a) shows that various estimates of standard error are related to bootstrap estimates. Estimates can also be obtained based on influence function estimates. Generally, influence functions for parameters are functional derivatives that can be expressed in terms of functionals of a distribution. Basically, they determine the influence of a single observation on an estimate, as a function of its value (location in the sample space). Since in practice the underlying distribution is unknown, the empirical distribution is often used as a plug-in estimate to the formula for the influence

function to give a sample estimate of the influence and is called the empirical influence function.

Using the empirical influence function, Efron shows that the influence function estimate of standard error is the same as the one obtained using Jaeckel's infinitesimal jackknife Efron (1982a, p. 42). Amari (1985) provides a thorough treatment of related differential geometry methods.

Following Efron (1982a, pp. 39–42), the infinitesimal jackknife estimate of the standard error is defined as follows:

$$\mathrm{SD}_{\mathrm{IJ}}(\theta_e) = \left(\frac{\sum_1^n U_i^2}{n^2} \right)^{1/2}$$

where θ_e is the estimate of the parameter θ and U_i is a directional derivative in the direction of the ith coordinate centered at the empirical distribution function.

Slight differences in the choice of the influence function estimate can lead to different estimates (e.g., the ordinary jackknife and the positive jackknife). See Efron (1982a, p. 42) for details.

The key result relating these jackknife and influence function estimates to the delta method is Theorem 6.2 of Efron (1982a, p. 43), which states that the nonparametric delta method and the infinitesimal jackknife give identical estimates of the standard error of an estimator in cases where the nonparametric delta method is defined.

We have seen that in the context of estimating standard errors, the jackknife, the bootstrap, and the delta methods are closely related and, in fact, are asymptotically equivalent. For the practitioner, however, their differences in small-to-moderate sample sizes is important.

General conclusions are that the bootstrap tends to be superior, but the bootstrap requires the use of Monte Carlo replications. The ordinary jackknife is second best. Too often, the nonparametric delta method (or, equivalently, the infinitesimal jackknife) badly underestimates the true standard errors.

Hall (1992a, pp. 86–88) discusses a slightly more general version of the delta method. He assumes that S_n and T_n are two asymptotically normal statistics that admit an Edgeworth expansion. If the two statistics differ by an amount that goes to zero in probability at a rate of $n^{-j/2}$ for $j \geq 1$, then the Edgeworth expansions of their distribution functions will differ by no more than $n^{-j/2}$. The standard delta method that we have described in this section amounts to the special case where S_n is a linear approximation to T_n. That is what we get by truncating the Taylor series expansion of φ with only the linear term.

Hall (1992a) goes on to point out the usefulness of this more general delta method. It may be easier to derive the low-order terms of the Edgeworth expansion for S_n rather than for T_n. Because S_n and T_n are "close," their Edgeworth expansions can only differ in the terms of order $n^{-k/2}$, where $k \geq j$. It may be sufficient to only obtain the expansion up to $n^{-(j-1)/2}$, in which case the

delta method is a convenient tool. Note that here we are using the delta method as an analytical device for obtaining Edgeworth expansion terms and not as an estimator per se.

6.1.3. Cross-Validation

Cross-validation is a general procedure used in statistical model building. It can be used to decide on the order of a statistical model (including time series models, regression models, mixture distribution models, and discrimination models). It also has been generalized to estimate smoothing parameters in nonparametric density estimation and to construct spline functions. As such, it is a very useful tool. The bootstrap provides a competitor to cross-validation in all such problems. The research on the bootstrap has not developed to the point where clear guidelines can be given for each of these problems.

The basic idea behind cross-validation is to take two random subsets of the data. Models are fit or various statistical procedures are applied to the first subset and then are tested on the second subset. The extreme case of fitting to all but one observation and then testing on the remaining one is sometimes referred to as leave-one-out and has also been called cross-validation by Efron because it is an often preferred special case of the general form of cross-validation.

Since leaving only one observation out does not provide an adequate test, the procedure actually is to fit the model n times, each time leaving out a different observation and testing the model on estimating or predicting the observation left out each time. This provides a fair test by always testing on observations not used in the fit. It also is efficient in the use of the data for fitting the model since $n - 1$ observations are always used in the fit.

In the context of estimating the error rate of linear discriminant functions (Section 2.1.2) we found that the bootstrap and variants (particularly the .632 estimate) were superior to leave-one-out in terms of mean square estimation error.

For classification trees (i.e., discriminant rules based on a series of binary decisions graphically represented in a tree structure), Breiman, Friedman, Olshen, and Stone (1984) use cross-validation to "prune" (i.e., remove or shorter branches) classification tree algorithms. They also discuss a bootstrap approach (pp. 311–313). They refer to Efron (1983) for the discriminant analysis example of advantages of the bootstrap over cross-validation but did not have theory or simulation studies to support its use in the case of classification trees. Further work has been done, but nothing has shown strong superiority for bootstrap.

6.1.4. Subsampling

Subsampling methods go back to Hartigan (1969), who developed the theory of confidence intervals for random subsampling through the typical value

theorem when the estimate is an *M*-estimator. As we saw in Chapter 3, Hartigan's results motivated Efron to propose his bootstrap percentile method for confidence intervals. Subsampling has subsequently been further developed by several authors, some motivated by early developments on the bootstrap in the 1980s. Subsampling has been applied for confidence intervals and variance estimates in both IID and dependent situations.

Politis, Romano, and Wolf (1999) summarize results on subsampling and compare it to the bootstrap. They include applications to IID samples, stationary and nonstationary time series, random fields, and marked point processes. The book includes both theory and simulation. In Chapter 2 they establish that subsampling methods converge at a first-order rate under weaker conditions than have been established for the bootstrap. In addition to the book, Politis and Romano have both made major contributions to the theory of bootstrap and subsampling.

Section 2.8 in Lahiri (2003a) provides a brief summary of results on subsampling methods. In Section 2.7, Lahiri (2003a) describes methods called generalized block bootstrap (GBB), of which the circular block bootstrap (CBB) and the stationary bootstrap (SB) or stationary block bootstrap are particular examples. Politis and Romano (1992a) developed CBB and Politis and Romano (1994a) introduced SB.

Lahiri (2003a) points out that subsampling is a moving block bootstrap (MBB) with block size equal to 1. He then goes on to define a generalized subsampling method that is very similar to GBB. See Lahiri (2003a, p. 39) for details.

As a guideline for usage of subsampling, there is a subsampling approach that works whenever the bootstrap works. The subsampling method only has first-order accuracy, whereas bootstrap estimates are often second-order accurate. On the other hand, requirements for consistency are weaker than for the bootstrap and so they can safely be applied in some situations where the smoothness conditions used to show consistency for the bootstrap do not hold.

Hartigan (1969, 1975) were the pioneering articles on subsampling. The monograph of Efron (1982a) reminded researchers about subsampling and may have motivated later research which includes Carlstein (1986), Politis and Romano (1992a, 1993a, 1994a–c), Hall and Jing (1996), and Bickel, Götze, and van Zwet (1997).

6.2. BOOTSTRAP VARIANTS

In previous chapters, we have introduced some modifications to the "nonparametric" bootstrap (i.e., sampling with replacement from the empirical distribution). These modifications were found sometimes to provide improvements over the nonparametric bootstrap when the sample size is small.

Recall that for the error rate estimation problem the .632 estimator, the double bootstrap (a form of bootstrap iteration), and the "convex" bootstrap

(a form of smoothing the bootstrap distribution) were variations that proved to be superior to the original nonparametric bootstrap in a variety of small sample simulation studies.

For confidence intervals, Hall has shown that the accuracy of both kinds of bootstrap percentile methods can be improved by bootstrap iteration. Efron's bias correction with the acceleration constant also provides a way to improve on the accuracy of his version of the bootstrap percentile confidence intervals.

When the problem indicates that the observed data should be modeled as coming from a distribution that is continuous and has a probability density, it may be reasonable to replace the empirical distribution function with a smoothed version (possibly based on kernel methods). This is referred to as the smoothed bootstrap.

Although it is desirable to smooth the distribution, particularly when the sample size is small, there is a catch. Kernel methods generally require large samples, particularly to estimate the tails of the density.

There is also the question of determining the width of the kernel (i.e., the degree of smoothing). Generally, kernel widths have been determined by cross-validation. Therefore, it is not clear whether or not there will be a payoff to using a smoothed bootstrap, even when we know that the density exists. This issue is clearly addressed by Silverman and Young (1987).

In fact, we may look at the bootstrap as another approach to deciding on the width of a kernel in density estimation (as a competitor to cross-validation). To this point I have not seen any comparisons of the bootstrap and cross-validation with respect to kernel density estimation.

Another variation on the bootstrap is Rubin's Bayesian bootstrap [see Rubin (1981)]. The Bayesian bootstrap can be viewed as a Bayesian's justification for using bootstrap methods as Efron and others have interpreted it. On the other hand, Rubin used it to point out weaknesses in the original nonparametric version of the bootstrap. Ironically, these days there are a number of interesting applications of the Bayesian bootstrap, particularly in the context of missing data [see Lavori et al. (1995) and Chapter 12 in Dmitrienko, Chuang – Stein, and D'Agostino (2007)].

6.2.1. Bayesian Bootstrap

Consider the case where $x_1, x_2, \ldots, x_n$ can be viewed as a sample of n independent identically distributed realizations of random variables $X_1, X_2, \ldots, X_n$, each with distribution F, and denote the empirical distribution by $\hat{F}$. Recall that the nonparametric bootstrap samples with replacement from $\hat{F}$. Let be a parameter of the distribution F. For simplicity we may think of x_i as one-dimensional and as a single parameter, but both could be multidimensional as well. Let ϕ be an estimate of based on $x_1, x_2, \ldots, x_n$. As we know, the nonparametric bootstrap can be used to approximate the distribution of $\hat{\varphi}$.

Instead of sampling each x_i with replacement and with probability $1/n$, the Bayesian bootstrap uses a posterior probability distribution for the X_i's. This posterior probability distribution is centered at $1/n$ for each X_i, but varies from one Bayesian bootstrap replication to another. Specifically, the Bayesian bootstrap replications are defined as follows: Draw $n-1$ uniform random variables from the interval [0, 1]. Let $u_{(1)}, u_{(2)}, \ldots, u_{(n-1)}$ denote their values in increasing order. Let $u_{(0)} = 0$ and $u_{(n)} = 1$. Then define $g_i = u_{(i)} - u_{(i-1)}$ for $i = 1, 2, 3, \ldots, n$. Then the g_i's are called the gaps between uniform order statistics. The vector

$$g = \begin{pmatrix} g_1 \\ g_2 \\ . \\ . \\ . \\ g_n \end{pmatrix}$$

is used to assign probabilities to the Bayesian bootstrap sample. Namely, n observations are selected by sampling with replacement from $x_1, x_2, \ldots, x_n$, but instead of each x_i having exactly probability $1/n$ of being selected each time, x_1 is selected with probability g_1, x_2 with probability g_2, and so on.

A second Bayesian bootstrap replication is generated in the same way, but with a new set of $n-1$ uniform random numbers and hence a new set of g_i's.

It is Rubin's point that the bootstrap and the Bayesian bootstrap are very similar and have common properties. Consequently, he suggests that any limitations attributable to the Bayesian bootstrap may be viewed as limitations of the nonparametric bootstrap as well. An advantage to a Bayesian is that it can be used to make the usual Bayesian-type inferences about the parameter ϕ based on ϕ's estimated posterior distribution, whereas, strictly speaking, the nonparametric bootstrap has only the usual frequentist's interpretation about the distribution of the statistic $\hat{\varphi}$.

If we let $g_i^{(1)}$ be the value of g_i in the first Bayesian bootstrap replication and let $g_i^{(2)}$ be the value of g_i in the second replication, we find based on elementary results for uniform-order statistics [see David (1981), for example] that

$$E(g_i^{(1)}) = E(g_i^{(2)}) = 1/n, \quad \mathrm{Var}(g_i^{(1)}) = \mathrm{Var}(g_i^{(2)}) = (n-1)/n^3,$$
$$\text{and} \quad C(g_i^{(1)}, g_j^{(2)}) = C(g_i^{(2)}, g_j^{(2)}) = -1/(n-1),$$

where $E(\bullet)$, $\mathrm{Var}(\bullet)$ and $C(\bullet,\bullet)$ denote expectation, variance, and correlation over the respective replications. Because of the above properties, the

bootstrap distribution for $\hat{\varphi}$ and the Bayesian bootstrap posterior distribution for φ will be very similar in many applications.

Rubin (1981) provides some examples and shows that the Bayesian bootstrap procedure leads to a posterior distribution for ϕ which is Dirichlet and is based on a conjugate Dirichlet prior distribution. Rubin then goes on to criticize the Bayesian bootstrap because of the odd prior distribution that is implied.

He sees the Bayesian bootstrap as being appropriate in some problems but views the prior as restrictive and hence does not recommend it as a general inference tool. In situations where Rubin is uncomfortable with the Bayesian bootstrap, he is equally uncomfortable with the nonparametric bootstrap. His main point is that through the analogy he makes with the nonparametric bootstrap, the nonparametric bootstrap also should not be oversold as a general inference tool. Much of the criticism is directed at the lack of smoothness of the empirical distribution. Versions such as the parametric bootstrap and the smoothed bootstrap overcome some of these objections. See Rubin (1981) for a more detailed discussion along with some examples.

The Bayesian bootstrap can be generalized by not restricting the prior distribution to be Dirichlet. The generalized version can be viewed as a Monte Carlo approximation to a posterior distribution for φ. In recent years there have been a number of papers written on the Bayesian bootstrap. Consult the bibliography for more references [particularly Rubin and Schenker (1998)].

6.2.2. The Smoothed Bootstrap

One motivation for the nonparametric bootstrap is that $\hat{F}$ (the empirical distribution) is the maximum likelihood estimator of F when no assumptions are made about F. Consequently, we can view the bootstrap estimates of parameters of F as nonparametric maximum likelihood estimates of those parameters.

However, in many applications, it is quite sensible to consider replacing $\hat{F}$ by a smooth distribution based on, say, a kernel density estimate of F' (i.e., the derivative of F with respect to x in the case of a univariate distribution F). The Bayesian version of this is given in Banks (1988).

Efron (1982a) illustrates the application of smoothed bootstrap versions for the correlation coefficient using Gaussian and uniform kernel functions. The observations in his simulation study were Gaussian, and the results show that the smoothed bootstrap does a little better than the original nonparametric bootstrap in estimating the standard error of the correlation coefficient.

Although smoothed versions of the bootstrap were considered early on in the history of the bootstrap, some researchers have recently proposed a Monte Carlo approximation based on sampling from a kernel estimate or a parametric estimate of F and have called it a generalized bootstrap [e.g., Dudewicz (1992)].

In his proposed generalized bootstrap, Dudewicz suggests fitting the observed data to a broad class of distributions and then doing the resampling from the fitted distribution. One such family that he suggests is called the *generalized lambda distribution* [see Dudewicz (1992, p. 35)].

This generalized lambda distribution is a four-parameter family that can be specified by a mean, variance, skewness, and kurtosis. The method of moments is a suggested estimation method that specifies the distribution by applying the sample mean, variance, skewness, and kurtosis to the corresponding population parameter in the distribution.

Comparisons of generalized bootstrap with the nonparametric bootstrap in a particular application are given by Sun and Muller-Schwarze (1996). They apply the generalized lambda distribution.

It appears that the generalized bootstrap might be a promising alternative to the nonparametric bootstrap since it has the advantage of taking account of the fact that the data come from a continuous distribution, but it does not seem to suffer any of the drawbacks of the smoothed bootstrap. Another technique also referred to as a generalized bootstrap is presented in Bedrick and Hill (1992).

The value of a smoothed bootstrap is not altogether clear. It depends on the context of the problem and the sample size. See Silverman and Young (1987) for more discussion of this issue.

6.2.3. The Parametric Bootstrap

Efron (1982a) views the original bootstrap as a nonparametric maximum likelihood approach. As such, it can be viewed as a generalization of Fisher's maximum likelihood approach to the nonparametric framework. When looked at this way, $\hat{F}$ is the nonparametric estimate of F. If we make no further assumptions, the ordinary nonparametric bootstrap estimates are "maximum likelihood."

If we assume further that F is absolutely continuous, then smoothed distributions are natural and we are led to the smoothed bootstrap. Taking this a step further, if we assume that F has a parametric form such as, say, the Gaussian distribution, then the appropriate estimator for F would be a Gaussian distribution with the maximum likelihood estimates of μ and σ^2 used for these respective unknown parameters. Sampling with replacement from such a parametric estimate of F leads to bootstrap estimates that are maximum likelihood estimates in accordance with Fisher's theory. The Monte Carlo approximation to the parametric bootstrap is simply an approximation to the maximum likelihood estimate.

The parametric bootstrap is discussed briefly on pp. 29–30 of Efron (1982a). It is interesting to note that a parametric form of bootstrapping is equivalent to maximum likelihood. However, in parametric problems, the existing theory on maximum likelihood estimation is adequate and the bootstrap adds little or nothing to the theory. Consequently, it is uncommon to see the parametric

bootstrap used in real problems. In more complex problems there may be semiparametric approaches that the author might refer to as a parametric bootstrap.

Davison and Hinkley (1997) justify the nonparametric bootstrap in parametric situations as a check on the robustness and/or validity of the parametric method. They introduce the parametric bootstrap through an example of an exponential distribution and describe the implementation in a section on parametric simulation (pp. 15–21). There they justify the use of parametric simulation. They then justify the use of the parametric bootstrap in cases where the estimator of interest has a distribution that is difficult to derive analytically or has an asymptotic distribution that does not provide a good small sample approximation, particularly for the variance, which is where the bootstrap is often useful.

6.2.4. Double Bootstrap

The double bootstrap is a method originally suggested in Efron (1983) as a way to improve on the bootstrap bias correction of the apparent error rate of a linear discriminant rule. As such, it is the first application of bootstrap iteration (i.e., taking resamples from each bootstrap resample). We briefly discussed this application in Chapter 2.

Normally, bootstrap iteration requires a total of B^2 bootstrap samples where B is both the number of bootstrap replications from the original sample and the number of bootstrap samples taken from each bootstrap replication. In Efron (1983), a Monte Carlo swindle is used to obtain the accuracy of the B^2 bootstrap samples with just $2B$ samples.

Bootstrap iteration has been particularly useful in improving the accuracy of confidence intervals. The theory of bootstrap iteration for confidence intervals was developed by Hall, Beran, and Martin and is nicely summarized in Hall (1992a). See Chapter 3, Section 3.1.4 for more detail. In general, bootstrap iteration can occur more than once and the order of accuracy of the bootstrap estimate increases with each iteration. However, as noted when comparing the ordinary bootstrap with the double bootstrap, there is also a price paid in terms of increased computer intensity with each iteration.

6.2.5. The *m*-out-of-*n* Bootstrap

When the nonparametric bootstrap was first introduced Efron proposed taking the sample size in the bootstrap sample to be the same as the sample size n in the original sample. This seemed reasonable and worked quite well in many applications. However, it was clear to the early researchers that the choice of a bootstrap sample size could in general to be taken as $m < n$ where n is the sample size of the original sample. Such a bootstrap has been called the m out of n bootstrap. It has been studied by various authors and Bickel, Götze, and van Zwet (1997) study it in detail.

In Chapter 9 we will discuss a number of situations where the naive nonparametric bootstrap fails to be consistent, but an m-out-of-n bootstrap with m appropriately chosen provides a remedy by giving consistent estimates. Since the m-out-of-n bootstrap method is so basic and easy to describe, the result of using it to obtain consistency is often called a "quick fix." Usually the asymptotic theory requires $m \to \infty$ as $n \to \infty$, but at a slower rate such that $m/n \to 0$. Amazingly, such a simple remedy works in a large number of examples including both dependent and independent observations.

CHAPTER 7

Efficient and Effective Simulation

In Chapter 1 we introduced the notion of a Monte Carlo approximation to the bootstrap estimate of a parameter, θ. We also mentioned that the bootstrap folklore suggests that the number of Monte Carlo iterations should be on the order of 100–200 for estimates such as standard errors and bias but 1000 or more for confidence intervals. These rules of thumb are based mostly on simulation studies and experience with a wide variety of applications.

Efron (1987) presented an argment that showed, based on calculations for the coefficient of variation, the 100 bootstrap iterations are all that are actually needed for the estimation of standard errors and sometimes a mere 25 will suffice. He also argues in favor of 1000 iterations for to get good estimates for the endpoints of bootstrap confidence intervals.

Booth and Sarkar (1998) challenge Efron's argument. They claim that the number of bootstrap iterations should be based on the conditional distribution of the coefficient of variation estimate rather than the unconditional distribution. They argue that the number of iterations should be sufficiently large so that the Monte Carlo portion of the error in estimation is so small as to have no effect on the statistical inference. They suggest using a conditioning argument that 800 iterations are needed for standard errors as compared to the 100 recommended by Efron. Section 7.1 deals with this topic in detail.

A somewhat theoretical basis for the number of iterations has been developed by Hall using Edgeworth expansions. He also has results suggesting the potential gain from various variance reduction schemes, including the use of antithetic variates, importance sampling, linear approximations, and balanced sampling. Details can be found in Appendix II of Hall (1992a).

In this chapter (with Section 7.1 dealing with uniform resampling or ordinary Monte Carlo) we summarize Hall's finding and provide guidelines for practitioners based on current developments. In addition to Hall, a detailed

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

account of various approaches can be found in Davison and Hinkley (1997, Chapter 9).

We should point out that in the 1980s and 1990s the computing speed was increasing dramatically, making computer-intensive methods more and more practical. By the same token, looking at this in 2007 it is clear that simple simulations can be run 100,000 times or more in a matter of seconds. So the need for efficient simulation is not nearly as great today. Still the faster we simulate, the more inventive we become. For example, if we want to study the properties of Bayesian estimates obtained by Markov chain Monte Carlo methods, we might think of a way of imbedding the estimation process within a bootstrap. Iterating two computer-intensive methods together makes for a very intensive method that might still benefit from efficient simulation. So I do not think efficient simulation has become a dead issue.

7.1. HOW MANY REPLICATIONS?

The usual Monte Carlo method, sampling with probability $1/n$ for each observation and with replacement from the original sample of size n, is referred to as a uniform resampling in Hall (1992a), and we shall adopt that terminology here.

Let B be the number of bootstrap replications in a uniform resampling. Let σ_B^2 be the variance of a single bootstrap resample estimate of the parameter. Since the Monte Carlo approximation to the bootstrap estimate is an average of B such estimates independently drawn, the variance of the Monte Carlo approximation is just $B^{-1}\sigma_B^2$.

Of course, this basic result is well known and has been applied for many years to judge how many replications to take in a simulation. There is nothing new here with the bootstrap. For the bootstrap, the particular distribution being sampled is the empirical distribution, but otherwise nothing is different. If we have the parameter $\sigma = F(x)$, where F is the population distribution and x is a specified value, then we obtain $\sigma_B^2 = \sigma(1-\sigma)$. By substituting $\sigma(1 - \sigma)$ for the result above, we have $B^{-1}\theta(1 - \theta) < (4B)^{-1}$ since $\sigma(1 - \sigma)$ must be less than or equal to 1/4.

This result can be generalized slightly. Hall (1992a) points out that if the estimate $\hat{\theta}$ is for a parameter θ that has a distribution function evaluated at x or is a quantile of a distribution function, then the variance of the uniform bootstrap approximation is CB^{-1} for large n and B. The constant C does not depend on B or n but is a function of unknown parameters. Often C can be usefully bounded above such as with the value 1/4 given earlier.

The practitioner then chooses B to make the variance sufficiently small, ensuring that the bootstrap approximation is close to the actual bootstrap estimate. Note that the accuracy of this approximation depends on B and not n. It only expresses how close the approximation is to the bootstrap estimate and does not express how close the bootstrap estimate is to the true parameter value!

If the constant C cannot be easily estimated or bounded, consider the following practical guideline. The practitioner can take, say, 100 bootstrap resamples and then double it to 200 to see how much the bootstrap approximation changes. He continues until the change is small enough. With the speed now available with modern computers, this approach is practical and commonly used.

When the parameter is a smooth function of a population mean, the variance is approximately $CB^{-1}n^{-1}$. Variance reduction methods can be used to reduce this variance, by reducing either C or the factor involving n (e.g., changing it from n^{-1} to n^{-2}).

The rules of thumb described by Efron (1987) and Booth and Sarkar (1998) are based on mathematical results which indicate that after a particular number of iterations the error in the bootstrap estimate is dominated by the error due to using the empirical distribution to mimic the behavior of the true distribution with the error due to the Monte Carlo approximation being relatively small. Even in the late 1990s, it seemed silly to argue between 100 and 800 iterations when for simple problems it is easy to complete the bootstrapping using 5000 to 10,000 iterations. In 2007, rapid calculation with 100,000 or more bootstrap iterations is commonplace.

When applicable, Hall's result provides specific accuracy bounds based on the specified problem and the desired accuracy. So I prefer using it as opposed to either Efron's or Booth–Sarkar general rule of thumb.

7.2. VARIANCE REDUCTION METHODS

Variance reduction methods or swindles (as they are sometimes referred to in the statistics literature) are tricks that adjust the sampling procedure with the goal of reducing the variance for a fixed number of iterations. It is an old idea that goes back to nuclear applications in the 1950s.

Historically, one of the earliest examples is the method of antithetic variates, which can be attributed to Hammersley and Morton (1956). A good survey of these early methods is Hammersley and Handscomb (1964).

A principle that is used to reduce the variance is to split the computation of the estimate into deterministic and stochastic components and then to apply the Monte Carlo approximation only to the stochastic part. This approach is applicable in special isolated cases and does not have a name associated with it. It was used, for example, in the famous Princeton robustness study (Andrews, Bickel, Hampel, Huber, Rogers, and Tukey, 1972).

7.2.1. Linear Approximation

The linear approximation is a special case of the idea expressed in the preceding paragraph. The estimator is expressed as the expected value of a Taylor series expansion. Since the linear term in the series is known to have zero

expectation, the Monte Carlo method is applied only to the estimation of the higher-order terms.

As an example, consider the bias estimation problem described in Section 2.1. Recall that the bootstrap estimation of bias is $E(\theta^* - \hat{\theta})$, where θ^* is an estimate of θ based on a bootstrap sample. We further assume that $\hat{\sigma} = g(\bar{x})$, where $\bar{x}$ is a sample mean and g is a smooth function (i.e., g has first- and higher-order derivatives in its argument).

A Taylor series expansion for $\theta^* - \hat{\theta}$ is

$$U^* = \theta^* - \hat{\theta} = (\bar{x}^* - \bar{x})g'(\bar{x}) + \frac{1}{2}(\bar{x}^* - \bar{x})^2 g''(\bar{x}) + \cdots \tag{7.1}$$

Now $E(U^*|\bar{x})$ from (7.1) is

$$E(U^*|\bar{x}) = E(\bar{x}^* - \bar{x}|\bar{x})g'(\bar{x}) + \frac{1}{2}E((\bar{x}^* - \bar{x})^2|\bar{x})g''(\bar{x}) + \cdots \tag{7.2}$$

$E(U^*|\bar{x})$ is the bootstrap estimate of bias. Uniform sampling would estimate this directly without regard to the expansion.

But, $E(\bar{x}^*|\bar{x}) = \bar{x}$, since bootstrap sampling is sampling with replacement from the original sample. Therefore $E(\bar{x}^* - \bar{x}|\bar{x}) = 0$ and the first term (i.e., the linear term in the expansion of $E(U^*|\bar{x})$) can be omitted. So Eq. (7.2) reduces to

$$E(U^*|\bar{x}) = E(x^* - \bar{x}|\bar{x})g'(\bar{x}) + \frac{1}{2}E((x^* - \bar{x})^2|\bar{x})g''(\bar{x}) + \cdots \tag{7.3}$$

To take advantage of Eq. (7.3), we define $V^* = U^* - (\bar{x}^* - \bar{x})g'(\bar{x})$ and apply uniform sampling to V^* instead. In view of Eq. (7.3), $E(V^*|\bar{x}) = E(U^*|\bar{x})$. So averaging V^* approaches the bootstrap estimate as $B \to \infty$ just as U^* does, but we have removed a deterministic part $E(\bar{x}^* - \bar{x})g'(\bar{x})$, which we know equals zero.

In Hall (1992a, Appendix II), this result is shown for the more general case where $\bar{x}$ is a d-dimensional vector. He shows that $\mathrm{Var}(V^*|\bar{x})$ is of the order $B^{-1}n^{-2}$ as compared to $B^{-1}n^{-1}$ for $\mathrm{Var}(U^*|\bar{x})$.

In principle, if higher-order derivatives exist, we may remove these terms and compute an estimate that approaches the bootstrap at a rate $B^{-1}n^{-k}$, where k is the highest order of derivatives removed. It does, however, require computation of a linear form in the first k central sample moments.

This principle can also be applied through the use of what are called control variates. An estimator T^* is decomposed using the following identity: $T^* = C + (T^* - C)$, where C is a "control variate." Obviously any variable can be chosen to satisfy the identity, but C should be picked (1) to have high positive correlation with T^* and (2) so that its statistical properties are known

analytically. Then to determine the statistical properties of T^*, we can apply the Monte Carlo approximation to $T^* - C$ instead of T^*.

Because of the high positive correlation between T^* and C, the variable $T^* - C$ will have a much smaller variance than T^* itself. So this device (i.e., the use of the control variate, C) enables us to get more precision in estimating T^* by only applying the Monte Carlo approximation to $T^* - C$ and using knowledge of the statistical properties of C. More details with specific applications to bootstrap estimates of bias and variance can be found in Davison and Hinkley (1997, pp. 446–450).

7.2.2. Balanced Resampling

Balanced resampling was introduced by Davison, Hinkley, and Schechtman (1986). It is also covered with a number of illustrative examples in Davison and Hinkley (1997, pp. 438–446).

The idea is to control the number of times observations occur in the bootstrap samples so that in the B bootstrap samples, each observation occurs the same number of times (namely, B). Of course for the bootstrap to work, some observations must be missing in certain bootstrap samples, while others may occur two or more times.

Balanced resampling does not force each observation to occur once in each sample but equalizes the number of occurrences of each observation over the set of bootstrap samples. If an observation occurs twice in one bootstrap sample, there must be another bootstrap sample where it is missing from the sample. This is reminiscent of the kind of balancing constraints used in statistical experimental designs (e.g., balanced incomplete block designs).

A simple way to achieve balanced resampling is to create a string of the observations $X_1, X_2, \ldots, X_n$ repeated B times (i.e., we have the sequence $Y_1, Y_2, \ldots, Y_{Bn}$, where $Y_i = X_j$ with j being the remainder when dividing i by n. Then take a random permutation π of the integers from 1 to Bn. Take $Y_{\pi(1)}, Y_{\pi(2)}, \ldots, Y_{\pi(n)}$ as the first bootstrap sample $Y_{\pi(n+1)}$, take $Y_{\pi(n+2)}, \ldots, Y_{\pi(2n)}$ as the second bootstrap sample, and so on, until $Y_{\pi((B-1)n+1)}, Y_{\pi((B-1)n+2)}, \ldots, Y_{\pi(Bn)}$ is the Bth bootstrap sample.

Hall (1992a, Appendix II) shows that balanced resampling produces an estimate with conditional variance on the order of $B^{-1}n^{-2}$. His result applies to smooth functions of a sample mean. Balanced resampling can be applied in much greater generality including the estimation of distributions and quantiles. In such cases there is still an improvement in the variance but not as dramatic an improvement. Unfortunately for distribution functions, the order is still Cn^{-1} and only the constant C is reduced. See Hall (1992a, pp. 333–335) for details.

The MC estimator of Chernick, Murthy, and Nealy (1985), discussed in Section 2.2.2, is a form of controlled selection where an attempt is made to sample with the limiting repetition frequencies of the bootstrap distribution.

As such, it is similar to variance reduction methods like balanced resampling.

7.2.3. Antithetic Variates

As mentioned earlier, the concept of antithetic variates dates back to Hammersley and Morton (1956). The idea is to introduce negative correlation between pairs of Monte Carlo samples to reduce the variance.

The basis for the idea is as follows: Suppose that $\hat{\varphi}_1$, and $\hat{\varphi}_2$ are two unbiased estimates for the parameter, φ. We can then compute a third unbiased estimate,

$$\hat{\varphi}_3 = \frac{1}{2}(\hat{\varphi}_1 + \hat{\varphi}_2).$$

Then

$$\mathrm{Var}(\varphi_3) = \frac{1}{4}[\mathrm{Var}(\hat{\varphi}_1) + 2\mathrm{Cov}(\hat{\varphi}_1, \hat{\varphi}_2) + \mathrm{Var}(\hat{\varphi}_2)].$$

Assume, without loss of generality, that $\mathrm{Var}(\hat{\varphi}_2) > \mathrm{Var}(\hat{\varphi}_1)$; then if $\mathrm{Cov}(\hat{\varphi}_1, \hat{\varphi}_2) < 0$, we have the following inequality:

$$\mathrm{Var}(\hat{\varphi}_3) \leq \frac{1}{2}(\mathrm{Var}\hat{\varphi}_2).$$

So $\hat{\varphi}_3$ has a variance that is smaller than half of the larger of the variances of the two estimates. If $\mathrm{Var}(\hat{\varphi}_1)$ and $\mathrm{Var}(\hat{\varphi}_2)$ are nearly equal, then, roughly, we have guaranteed a reduction by about a factor of two for the variance of the estimate. The larger the negative correlation between $\hat{\varphi}_1$ and $\hat{\varphi}_2$, the greater is the reduction in variance.

One way to do antithetic resampling is to consider the permutation, π, that maps the largest X_i to the smallest, the second largest to the second smallest, and so on. Let the odd bootstrap samples be generated by uniform bootstrap resampling. The even bootstrap samples take $X_i^* = X_{\pi(k)}$ where $X_i^* = X_k$ for the preceding odd bootstrap sample.

The pairs of bootstrap samples generated in this way are negatively correlated because the permutation π maps the indices for the larger values to the smaller values. So if the first bootstrap sample tends to have higher-than-average values, the second will tend to be lower-than-average values. This provides the negative correlation between each consecutive even–odd pair that leads to a negative correlation between the estimate derived from the even samples and an estimate derived from the odd samples and thus a reduction in variance to the estimator obtained by averaging the two.

The estimates that we compute for the odd bootstrap samples and even bootstrap samples, we call U_1^* and U_2^* respectively. The antithetic resampling estimate is then

$$U^* = (U_1^* + U_2^*)/2.$$

Unfortunately, Hall (1992a) shows that antithetic resampling only reduces the variance by a constant factor and hence is not as good as balanced resampling or the linear approximation.

7.2.4. Importance Sampling

Importance sampling is an old variance reduction technique. Reference to it can be found in Hammersley and Handscomb (1964). One of the first to suggest its use in bootstrapping is Johns (1988).

Importance sampling (or resampling) is a useful tool when estimating the tails of the distribution function or for quantile estimation. It has limited value when estimating bias and variance. It is, however, useful when estimating the tails of a distribution function or the quantiles. So, for example, it can be used for hypothesis tesing problems where the estimate of a p-value for a test statistic is important.

The idea is to control the sampling so as to take more samples from the part of the distribution which is important to the particular estimation problem. For example, when estimating the extreme tails of a distribution (i.e., $1 - F(x)$ for very large x or $F(x)$, we need to observe values larger than x.

However, if the probability $1 - F(x)$ is very small, n must be extremely large to even observe values greater than x. Even for extremely large n and $1 - F_n(x) > 0$, where F_n is the empirical distribution, the number of observations greater than x in the sample will be small.

Importance resampling is an idea used to improve such estimates by including these observations more frequently in the bootstrap samples. Of course, any time the sampling distribution is distorted by such a procedure, an appropriate weighting scheme is required to ensure that the estimate is converging to the bootstrap estimate as B gets large.

Basically, it exploits the identity that for a parameter μ defined as

$$\mu = \int m(y)\, dG(y) = \int m(y)\{dG(y)/dH(y)\}\, dH(y).$$

This suggests sampling from H instead of G by using the weight $dG(y)/dH(y)$ for each value of $m(y)$ that is sampled.

This can work only if the support of H includes the support of G (i.e., $G(y) \neq 0 \rightarrow H(y) \neq 0$). A detailed description of importance sampling can be found in Davison and Hinkley (1997, pp. 450–466).

One can view importance resampling as a generalization of uniform resampling. In uniform resampling each X_i has probability $1/n$. In general, we can

define an importance resample by assigning probability p_i to X_i where the only restriction on the p_i's is that $p_i \geq 0$ for each i and $\sum_{i=1}^{n} p_i = 1$.

When $p_i = 1/n$ the jth bootstrap sample mean $\bar{X}_j^* = \frac{1}{n}\sum_{i=1}^{n} \bar{X}_i^*$ is an unbiased estimate of $\bar{X}$, and the Monte Carlo approximation

$$\bar{X}_B^* = \frac{1}{B}\sum_{i=1}^{B} \bar{X}_j^*$$

approaches $\bar{X}$ as $B \to \infty$. This is a desirable property that is lost if for some values of i, $p_i \neq 1/n$. However, since

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

define $\bar{X}_j^* = \sum_{i=1}^{n} \alpha_i X_i^*$, where α_i is chosen so that if $X_i^* = X_k$, then $\alpha_i = 1/np_k$ for $k = 1, 2, \ldots, n$.

This weighting guarantees that conditional on $X_1, X_2, X_3, \ldots, X_n$, we obtain $E(\bar{X}_j^*) = \bar{X}$. One can then look for values for the p_k's so that the variance of the estimator (in this case $\bar{X}_j^*$) is minimized.

We shall not go into details of deriving optimal p_k's for various estimation problems. Our advice to the practitioner is to only consider importance sampling for cases where the distribution function, a p-value or a quantile, is to be estimated or in the specia.

Hall (1992a, Appendix II) derives the appropriate importance sample for minimizing the variance of an estimate of a distribution function for a studentized asymptotically normal statistic. The interested reader should look there for more details. Other references on importance resampling are Johns (1988), Hinkley and Shi (1989), Hall (1991a), and Do and Hall (1991a,b).

A clever application of importance resampling is referred to as bootstrap recycling. This can be applied to an iterated bootstrap by repeated use of the importance sampling identity, Eq. (7.4). It has the advantage when a statistic of interest is complicated and costly to estimate, as in the case of a difficult optimization problem, such as Bayesian estimation problems, which require Markov chain Monte Carlo methods for computing an a posteriori estimate it is less complicated. Details along with application to bootstrap iteration can be found in Davison and Hinkley (1997, pp. 463–466).

7.2.5. Centering

Recall the linear approximation V^* to the bias estimation problem described in Section 7.2.1 along with the smooth function g defined in that section. Let

$$\bar{X}^*_* = B^{-1}\sum_{j=1}^{B} \bar{X}^*_j.$$

Let

$$\hat{X}^*_B = B^{-1}\sum_{j=1}^{B} g(\bar{X}^*_j) - g(\bar{X}^*_*).$$

Now

$$U^* = B^{-1}\sum_{j=1}^{B} g(\bar{X}^*_j) - g(\bar{X}).$$

We choose to center at $g(\bar{X}^*_*)$ instead of $g(\bar{X})$. This idea was introduced in Efron (1988). Hall (1992a, Appendix II) shows that it is essentially equivalent to linear approximation in its closeness to the bootstrap estimate.

These variance reduction techniques are particularly useful for complex problems where the estimates themselves might require intensive computing as in some of the examples in Chapter 8 and the Bayesian methods that use Markov chain Monte Carlo in Section 8.11. For simple problems that do not involve iterated bootstraps, generating 10,000 or more bootstrap samples using uniform resampling should not be a problem.

7.3. WHEN CAN MONTE CARLO BE AVOIDED?

In the nonparametric setting, we have shown in Chapter 3 several ways to obtain confidence intervals. Most approaches to bootstrap confidence intervals require adjustments for bias and skewness including Efron's BC_a intervals and Hall's bootstrap iteration technique. Each requires many bootstrap replications.

Of course, in parametric formulations without nuisance parameters, classical methods provide exact confidence intervals without any need for Monte Carlo. This is because pivotal quantities can be constructed whose probability distributions are known or can be derived. Because of the duality between hypothesis testing and confidence intervals, the same statement applies to hypothesis tests.

In a "semiparametric" setting where weak distributional assumptions are made (e.g., the existence of a few moments of the distribution), asymptotic expansions of the distribution of these quanitites (which may be asymptotically pivotal) can be used to obtain confidence intervals.

We have seen, for example in Section 3.1.4, that the asymptotic properties of bootstrap iteration can be derived from Cornish–Fisher expansions. These

results suggest that some approaches are more accurate because of their faster rate of convergence.

DiCiccio and Efron (1990, 1992), using properties of exponential families, construct confidence intervals with the accuracy of the bootstrap BC_a intervals but without any Monte Carlo. Edgeworth and Cornish–Fisher expansions can be used in certain problems. The difficulty, in practice, is that they sometimes require large samples to be sufficiently accurate, particularly when estimating the tails of the distribution.

The idea of re-centering and expansion in the neighborhood of a saddlepoint was first suggested by Daniels (1954) to provide good approximations to the distribution of the test quantity in the neighborhood of a point of interest (e.g., the tails of the distribution).

Field and Ronchetti (1990) apply this approach, which they refer to as small sample asymptotics in a number of cases. They claim that their approach works well in small samples and that it obtains the accuracy of the bootstrap confidence intervals without resampling. Detailed discussion of saddlepoint approximations can be found in Davison and Hinkley (1997, pp. 466–485).

Another, similar approach due to Hampel (1973) is also discussed in Field and Ronchetti (1990). A recent expository paper on saddlepoint approximations is Reid (1988).

Applications in Field and Ronchetti (1990) include estimation of a mean using the sample mean, robust location estimators including L-estimators and multivariate M-estimators. Confidence intervals in regression problems and connections with the bootstrap in the nonparametric setting are also considered by Field and Ronchetti (1990).

Also, the greatest promise with small sample asymptotics is the ability of high-speed computers to generate the estimates and their apparent high accuracy in small samples. Nevertheless, none of the major statistical packages include small sample asymptotic methods to date, and the theory to back up the empirical evidence of small sample accuracy requires further development.

7.4. HISTORICAL NOTES

Variance reduction methods for parametric simulation have a long history, and the information is scattered throughout the literature and in many disciplines. Some of the pioneering work came out of the nuclear industry in the 1940s and 1950s when computational methods were a real challenge.

There were no fast computers then! In fact the first vacuum tube computers were developed in the mid-1940s at Princeton, New Jersey and Aberdeen, Maryland, to aid the effort in World War II. John von Neumann was one of the key contributors to computing machine development.

Some discussions of these variance reduction methods can be found in various texts on Monte Carlo methods, such as Hammersley and Handscomb

(1964), Bratley, Fox, and Schrage (1987), Ripley (1987), Devorye (1986), Mooney (1997), and Niederreiter (1992). My own account of the historical developments can be found in a chapter on Monte Carlo methods in a compendium on risk analysis techniques, written for employees at the Army Materiel Systems Analysis Activity (Atzinger, Brooks, Chernick, Elsner, and Foster, 1972). An early clever application to the comparison of statistical estimators for robustness was the publication that resulted from the Princeton robustness study (Andrews, Bickel, Hampel, Rogers, and Tukey). They gave the technique the colorful name "swindle."

Balanced bootstrap simulation was first introduced by Davison, Hinkley, and Schechtman (1986). Monte Carlo estimates are mentioned at the beginning of this chapter. Obgonmwan (1985) proposes a slightly different method for achieving first-order balance. Graham, Hinkley, John, and Shi (1990) discuss ways to achieve second-order balance, and they provide connections to the classical experimental designs. A recent overview of balanced resampling based on the use of orthogonal multiarrays is Sitter (1998). Nigam and Rao (1996) develop balanced resampling for finite populations when applying simple random sampling or stratified random sampling with equal samples per strata. Do (1992) compares balanced and antithetic resampling methods in a simulation study.

The theoretical aspects of balanced resampling were investigated by Do and Hall (1991b). There are mathematical connections to number-theoretical methods of integration (Fang and Wang, 1994) and to Latin hypercube sampling (McKay, Beckman, and Conover, 1979; Stein, 1987; Owen, 1992).

Importance resampling was first suggested by Johns (1988). Hinkley and Shi (1989) applied it to the iterated bootstrap confidence intervals. Gigli (1994a) outlines its use in parametric simulation for regression and time series.

The large sample performance of importance resampling has been investigated by Do and Hall (1991a). Booth, Hall, and Wood (1993) describe algorithms for it. Gigli (1994b) provides an overview on resampling simulation techniques.

Linear approximations were used as control variates in bootstrap sampling by Davison, Hinkley, and Schechtman (1986). Efron (1990) took a different approach using the re-centered bias estimate and control variates in quantile estimation. Therneau (1983) and Hesterberg (1988) provide further discussion on control variate methods.

The technique of bootstrap recycling originated with Davison, Hinkley, and Worton (1992) and was derived independently by Newton and Geyer (1994). Properties of bootstrap recycling are discussed for a variety of applications in Ventura (1997).

Another approach to variance reduction is Richardson extrapolation, which was suggested by Bickel and Yahav (1988). Davison and Hinkley (1997) and Hall (1992a) both provide sections discussing variance reduction methods. Davison and Hinkley (1997) briefly refer to Richardson extrapolation in

Problem 22 of Section 9.7, page 494. Hall (1992a) mentions it only in passing as another approach.

General discussion of variance reduction techniques for bootstrapping appear in Hall (1989a, 1992c). Hall (1989b) deals with bootstrap applications of antithetic variate resampling.

Saddlepoint methods originated with Daniels (1954), and Reid (1988) reviews their use in statistical inference. Longer accounts can be found in Jensen (1992). Recent applications to bootstrapping the studentized mean are given in Daniels and Young (1991). Field and Ronchetti (1990) and Barndorff-Nielsen and Cox (1989) also deal with saddlepoint methods. Other related asymptotic results can be found in Barndorff-Nielsen and Cox (1994).

CHAPTER 8

Special Topics

This chapter deals with a variety of statistical problems. The common theme is the complex nature of the problems. In many cases, classical approaches require special assumptions or they provide incomplete or inadequate answers.

Although the bootstrap theory has not advanced to the stage of explaining how well it works on these complex problems, many researchers and practitioners see the bootstrap as a valuable tool when dealing with these difficult, but practical, problems.

8.1. SPATIAL DATA

8.1.1. Kriging

When monitoring air pollution in a given region, the government wants to control the levels of certain particulates. By setting up monitoring stations at particular locations, the levels of these particulates can be measured at various times.

An important practical question is, given measured levels of a particulate at a set of monitoring stations, What is the level of the particulate in the communities located between the monitoring stations? Spatial stochastic models are constructed to try to realistically represent the way the pollution levels change smoothly as we move from one place to another. The results are graphically represented with contour plots such as the ones in Figure 8.1.

Kriging is one technique for generating such contour plots. It has optimality properties under certain conditions. Due to the statistical nature of the procedure, there is, of course, uncertainty about the particulate level and hence also the constructed contours. The bootstrap provides one approach to

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

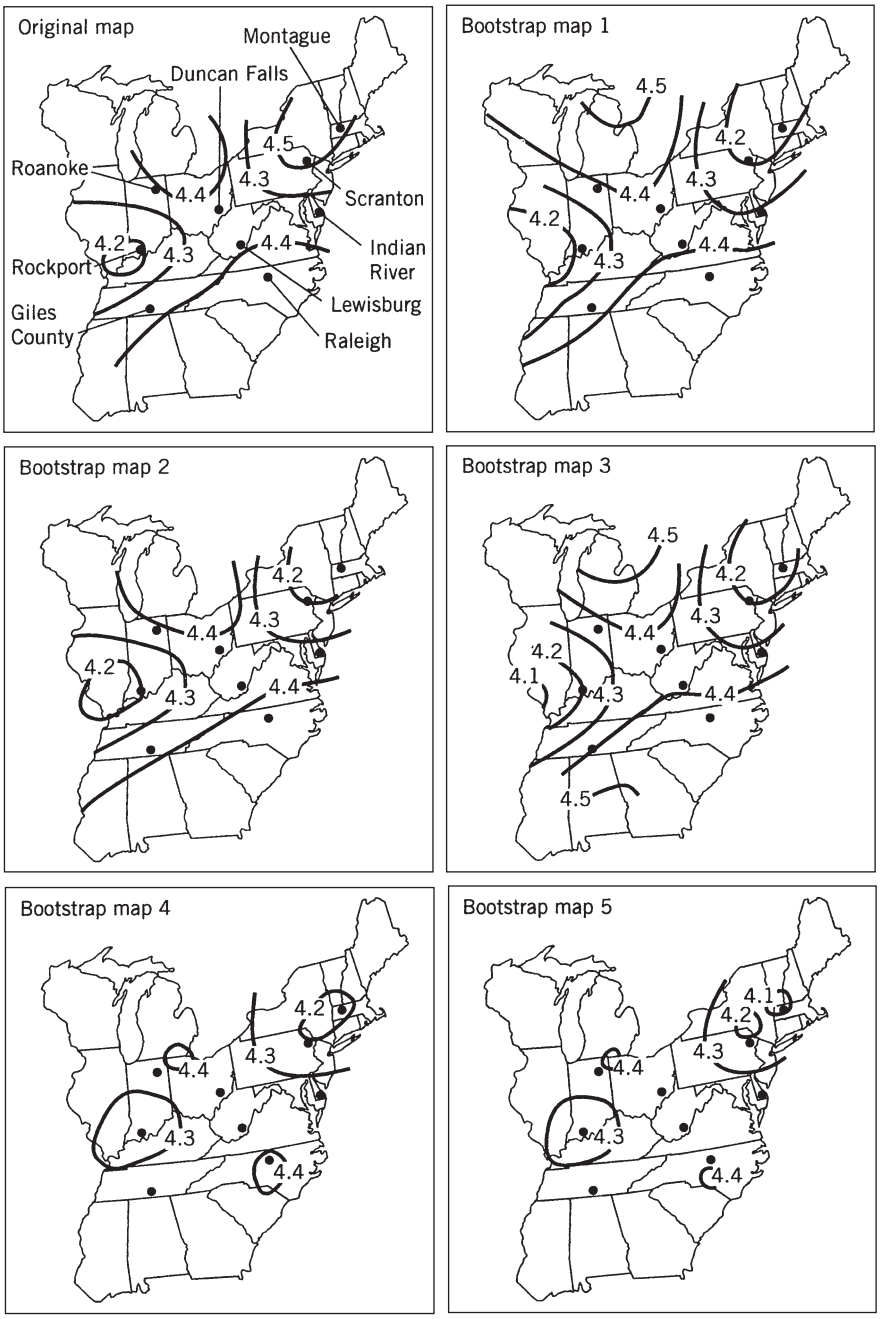


Figure 8.1 Original map and bootstrap maps. [From Diaconis and Efron (1983, p. 117), with permission from *Scientific American*.]

recognizing and representing that variability. As Diaconis and Efron (1983) show, bootstrap samples of kriging contours can be used as a visual tool that illustrates this variability.

Diaconis and Efron (1983) considered another important pollution problem that was extensively studied in the 1980s, namely the question of acidity levels in the rainfall, using 2000 measurements of pH levels for rainfall at nine weather stations in the Northeastern United States over a two-year period from September 1978 through August 1980. Figure 8.1 shows the original contour map generated by Switzer and Eynon, along with five maps generated from five bootstrap replications.

The weather stations are illustrated with nine dots, and the city names are labeled on the original map; there are also important noticeable differences. Although there is no generally accepted measure of the variability of contour lines on a map, visual inspection of the bootstrap replications provides a sense of this variability and points out that the original map must be interpreted cautiously.

For point estimates, we have confidence intervals to express our degree of uncertainty in the estimates. Similarly, we have uncertainty in the kriging contours. This problem is much more complex. The bootstrap procedure at least provides one way to assess this uncertainty.

A naive approach to bootstrapping spatial data is not appropriate because of the dependence structure as described through spatial correlation (Cressie 1991, pp. 489–497).

We shall not describe the kriging method here [see Cressie (1991) for details]. When Diaconis and Efron (1983) describe the weather maps, they do not mention kriging or the particular way the bootstrap was applied by Eynon and Switzer to obtain the bootstrap sample maps. Presumably, they bootstrapped the residuals from the model fit. This must have been done in a way so that the geographic relationship among the weather stations was preserved over the bootstrap replications.

Early work on bootstrapping spatial data that takes account of local (in terms of spatial distance) correlation is due to Hall (1985). He proposed two methods for bootstrapping. Both ideas start with the division of the entire space D into k congruent subregions $D_1, D_2, \ldots, D_k$.

Let $\mathbf{Z}_1\, \mathbf{Z}_2, \ldots, \mathbf{Z}_k$ be the k vectors of observations associated with the corresponding subregions. The first scheme assigns the bootstrap data $\mathbf{Z}_i^*$ for $i = 1, 2, 3, \ldots, k$, by sampling with replacement from the original data and if $\mathbf{Z}_i^* = \mathbf{Z}_j$, then the bootstrap observation is assigned to the region D_j. The second approach is to sample with replacement from the possible subregions and then the ith chosen subregion D_i^* has all its data $\mathbf{Z}_i$ assigned to the jth region D_j. if the jth region is selected on the ith draw. Now if $i \neq j$, then in the bootstrap sample the vector $\mathbf{Z}_k$ is now in a different region.

This latter scheme is similar to Kunsch's block bootstrap method (Künsch, 1989), except that the blocks correspond to special regions in Hall's case and to a subset of consecutive time points in Kunsch's case. Other semiparametric

approaches involve regression equations relating the spatial data to explanatory variables. These methods are described in Cressie (1991, pp. 493–494). The methods of Freedman and Peters (1984a) or Solow (1985) can then be exploited in the context of spatial data.

Cressie (1991, p. 496) describes a form of parametric bootstrap. He points out that the Monte Carlo based hypothesis testing approaches of Hope (1968) and Besag and Diggle (1977) predated Efron's celebrated *Annals of Statistics* paper (Efron, 1979a), and each was based on a suggestion from Barnard (1963). Cressie does not.

Stein (1999) is a text that covers the existing theory on kriging. There is no coverage of the bootstrap in it, but it is the authoritative reference on kriging.

8.1.2. Block Bootstrap on Regular Grids

In order to study the properties of resampling methods for randomly generated spatial data, we need to know the sampling mechanism that generates the original data because it is that mechanism that the resampling procedure must mimic. Different sampling mechanisms could lead to different results. Also to study asymptotic properties of the estimates we need to define a framework for the asymptotics in a spatial data setting. In this setting we will consider an extension of the block bootstrap from time series to spatial data.

The main difference between spatial and time series data is that in time series there is a natural direction of order (increase in time), but spatial data are two-dimensional or higher and in multidimensional Euclidean space there is no natural direction to define an ordering. So with time series the idea of asymptotics is naturally determined by t approaching infinity. But this is not the case with spatial data.

So with spatial data the choice of the concept "approaching infinity" can be defined but is not unique. These concepts are well covered in Cressie (1991, 1993). There are two basic paradigms that determine the various asymptotic structures. One paradigm is called the increasing domain asymptotics and the other is called infill asymptotics. For detailed coverage of these paradigms see Chapter 5 of Cressie (1991, 1993).

When the sampling sites are separated by a fixed distance and the sample region becomes unbounded as the number of sample sites increases, we have a form of increasing domain asymptotics. This is the most common framework for spatial data. It often leads to conclusions that are similar to what we have seen for time series. An example would be a process observed over increasing nested rectangles on a regular grid of integer spacing.

On the other hand, the infill asymptotics comes about when sites are all located in a fixed bounded region and the number of sites within that region increases to infinity. In this case as the sample size increases, the distance from

one site to its nearest neighbor tends to zero. This type of asymptotic theory is common with mining data and other geostatistical applications.

It is common knowledge in the realm of spatial statistics that infill asymptotics can lead to far different inferences than with the increasing domain asymptotics. The following references deal with this issue: Morris and Ebey (1984), Stein (1987b, 1989), Cressie (1993), and Lahiri (1996).

Another possibility is a combination of the two paradigms. One such case is called mixed increasing domain asymptotic structure. In this case the sample region grows without bounds. At the same time, the space between neighboring sample sites goes to zero as the number of samples grows.

To formally define the spatial block bootstrap requires a lot of machinery and term peculiar to spatial processes. Instead of doing that, I will describe how the blocks are generated in the two-dimensional case. In this section, we cover the consistency results for regular grids. Lahiri (2003a) describes a numerical example to illustrate the technique. These consistency results are also given in Lahiri (2003a). In the next section, we follow the results of Lahiri (2003a) with regard to irregular grids.

The regular grids in two-dimensions are rectangles. Depending on the shape of the sampling regions, the rectangles may or may not be squares. We take the blocks to overlap and be complete. By complete we mean that the collection of grids covers each subregion.

Details can be found in Lahiri (2003a). He is able to show that under certain conditions the spatial block bootstrap provides a consistent estimate for the variance [Lahiri, 2003a, Theorem 12.1, p. 295]. The theorem requires the assumption that the spatial process is a stationary random field satisfying a strong mixing condition. It is also possible to bootstrap the empirical distribution function to get a consistent estimate of the bootstrap under similar assumptions as for the variance.

8.1.3. Block Bootstrap on Irregular Grids

Lahiri defines classes of spatial designs that lead to irregularly spaced grids. He then proves consistency of bootstrap estimates that are based on these irregular grids (Lahiri, 2003a, Theorem 12.6). Under the spatial stochastic designs defined in Section 12.5.1 of Lahiri (2003a), the irregularly spaced grids are constructed and used in the consistency theorems. The results are analogous to the earlier results for long-range dependence in time series.

8.2. SUBSET SELECTION

In both regression and classification problems, we may have a large set of variables that we think can help to predict the outcome. In the case of regression, we are trying to predict an "average" response (usually a conditional

expectation); and in discriminant analysis, we are trying to predict the appropriate category to which the observation belongs.

In either case a subset of the candidate variable may provide most of the key information. Often when there are too many candidate variables, it will actually be better for prediction purposes to use a subset. This is often true because (1) some of the variables may be highly correlated or (2) we do not have a sufficiently large enough sample size to estimate all of the coefficients accurately.

To overcome these difficulties, there have been a number of procedures developed to help pick the "best" subset. For both discrimination and regression there are forward, backward, and stepwise selection procedures based on the use of criteria such as an F ratio, which compares two hypotheses at each stage.

Although there are many useful criteria for optimizing the choice of the subset, in problems with a large number of variables, the search for the "optimal" subset is unfeasible. That is why suboptimal selection procedures such as forward, backward, and stepwise selection are used. For a given data set, forward, backward, and stepwise selection may lead to different answers.

This should suggest to the practitioner that here is not a unique answer to the problem (i.e., different sets of variables may work equally well). Unfortunately, there is a great temptation for the practitioner to interpret the variables selected as being useful and those left out as not being useful. Such an interpretation can be a big mistake.

Diaconis and Efron (1983) relate an example of a study by Gong which addresses this issue through the bootstrap [see Efron and Gong (1983, Section 10) or Gong (1986) for more details]. In the example, a group of 155 people with acute and chronic hepatitis were initially studied by Dr. Gregory at the Stanford University School of Medicine. Of the 155 patients, 33 died from the disease and 122 survived.

Gregory wanted to develop a model to predict a patient's chance of survival. He had 19 variables available to help with the prediction. The question is, Which variables should be used? With only 155 observations, all 19 is probably too many. Using his medical judgment, Gregory eliminated six of the variables. Forward logistic regression was then used to pick 4 variables from the remaining 13.

Gong's approach was to bootstrap the entire data analysis including the preliminary screening process. The interesting results were in the variability of the final variables chosen by the forward logistic regression procedure. Gong generated 500 bootstrap replications of the data for the 155 patients. In some replications, only one predictor variable emerged. It could be ascites (the presence of a fluid in the abdomen), the concentration of bilirubin in the liver, or the physician's prognosis.

Other bootstrap samples led to the selection of as many as six of the variables. None of the variables emerged as very important by appearing in 60%

or more of the cases. In addition to showing that caution must be exercised when interpreting the results of a variable selection procedure, this example illustrates the potential of the bootstrap as a tool for assessing the effects of preliminary "data mining" (or exploratory analysis) on the final results. In the past, such effects have been ignored because they are difficult or impossible to assess mathematically.

McQuarrie and Tsai (1998) consider the use of cross-validation and bootstrap for model selection. They devote 40 pages to the comparison of these two resampling methods (see pp. 251–291). Their general conclusion was that the two procedures are equally good at selecting an appropriate model. For the bootstrap they considered both the vector and the residual approaches. Again, the comparison showed little difference because it is easier to bootstrap residuals in the model selection process, and that is why they recommend bootrapping residuals.

Data mining is becoming very popular for businesses with large data sets. People believe that intensive exploratory analysis could reveal patterns in the data that can be used to gain a competitive advantage. Consequently, it is now a growing discipline among computer scientists. Statistics and possibly the bootstrap could provide a helpful role in understanding and evaluating data mining procedures, particularly apparent patterns that could have occurred merely by chance and are therefore not really significant.

8.3. DETERMINING THE NUMBER OF DISTRIBUTIONS IN A MIXTURE MODEL

Mixture distribution problems arise in a variety of circumstances. In the case of Gaussian mixtures, they have been used to identify outliers [see Aitkin and Tunnicliffe Wilson (1980)] and to investigate the robustness of certain statistics for departures from normality [e.g., the sample correlation coefficient studied by Srivastava and Lee (1984)].

Increasing use of mixture models can be found in the field of cluster analysis. Various examples of such applications can be found in the papers by Basford and McLachlan (1985a–d).

In my experience, working on defense problems, targets, and decoys appeared to have feature distributions that fit well to a mixture of multivariate normal distributions. This could occur because the enemy could use different types of decoys or could apply materials to the targets to try to hide their signatures.

Discrimination algorithms would work well once the appropriate mixture distributions are determined. The most difficult part of the problem is to determine the number of distributions in the mixture for the targets and the decoys.

McLachlan and Basford (1988) apply a mixture likelihood approach to the clustering problem. As one approach to deciding on the number of

distributions in the mixture, they apply a likelihood ratio test employing bootstrap sampling.

The bootstrap is used because in most parametric mixture problems if we define λ to be the likelihood ratio statistic, $-2\log\lambda$ fails to have an asymptotic chi-square distribution because the regularity conditions necessary for this standard asymptotic result fail to hold. The bootstrap is used to approximate the distribution of $-2\log\lambda$ under the null hypothesis.

We now formulate the problem of likelihood estimation for a mixture model and then define the mixture likelihood approach. This will then be followed by the bootstrap test for the number of distributions.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be p-dimensional random vectors. Each $\mathbf{X}j$ comes from a multivariate distribution G which is a mixture of a finite number of probability distributions, say k, where $F_1, F_2, \ldots, F_k$ represent these distributions and π_1, $\pi_2, \ldots, \pi_k$ represent their proportions where π_j is the probability that population j is selected. Consequently, we require

$$\sum_{i=1}^{k} \pi_i = 1 \quad \text{and} \quad \pi_i \geq 0 \qquad \text{for } i = 1, 2, \ldots, k.$$

From this definition we also see that the cumulative distribution G is related to the F_i distributions by

$$G(\mathbf{x}) = P[\mathbf{X}_j \leq \mathbf{x}] = \sum_{i=1}^{k} \pi_i F_i(x),$$

where by $\mathbf{u} \leq \mathbf{v}$ we mean that each component of the p-dimensional vector $\mathbf{u}$ is less than or equal to the corresponding component of the p-dimension vector $\mathbf{v}$.

Assuming that G and the F_is are all differentiable (i.e., each has a well-defined density function), then if $g_\varphi(\mathbf{x})$ is the density for G and $f_{i\varphi}(\boldsymbol{x})$ is the density for $\boldsymbol{F}_i$, we obtain

$$g_\varphi(\mathbf{x}) = \sum_{i=1}^{k} \pi_1 f_{i,\phi}(x).$$

Here φ denotes the vector $\begin{pmatrix} \pi \\ \theta \end{pmatrix}$, where

$$\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_k \end{pmatrix}$$

and θ is a vector of unknown parameters defining the distributions $\boldsymbol{F}_i$ for $i = 1, 2, \ldots, k$. If we assume that all the f_i belong to the same parametric family of distributions, there will be an identifiability problem with the mixtures, since any rearrangement of the indices will not change the likelihood. One way to overcome this problem is to define the ordering so that $\pi_1 \geq \pi_2 \geq \pi_3 \geq \ldots \geq \pi_k$. Once this ambiguity is overcome, maximum likelihood estimates of the parameters can be obtained by solving the system of equations

$$\partial L\phi/\partial\phi = 0, \tag{8.1}$$

where L is the log of the likelihood function and φ is the vector of parameters.

The method of solution is the EM algorithm of Dempster, Laird and Rubin (1977) which was applied earlier to specific mixture models by Hasselblad (1966, 1969), Wolfe (1967, 1970) and Day (1969). They recognized through manipulation of Eq. (8.1) above that if we define the a posteriori probabilities

$$\tau_{i_j}(\phi) = \tau_j(x_j;\phi) = \text{probability that } x_j \text{ comes from } F_i$$

$$= \pi_i f_{i,\theta}(x_j) \Big/ \sum_{t=1}^{k} \pi_\ell f_{\ell,\theta}(x_j) \qquad \text{for } i = 1, 2, \ldots, k,$$

then

$$\hat{\pi}_i = \sum_{j=1}^{n} \hat{\tau}_{ij}/n \qquad \text{for } i = 1, 2, \ldots, k$$

and

$$\sum_{i=1}^{k}\sum_{j=1}^{n} \hat{\tau}_{ij}\partial \log f_{i\hat{\theta}}(x_j)/\partial\hat{\theta} = 0.$$

There are many issues related to the successful application of the EM algorithm in various parametric problems including the choice of starting values. For details, the reader should consult McLachlan and Basford (1988) or McLachlan and Krishnan (1997), a text dedicated to the EM algorithm and its applications.

The approach taken in McLachlan and Basford (1988), when choosing the number of distributions is to compute the generalized likelihood ratio statistics, which compare the likelihood under the null and alternative hypotheses. The likelihood ratio λ is then used to accept or reject the null hypothesis. As in classical theory, the quantity $-2\log\lambda$ is used for the test, but since the usual

regularity conditions are not satisfied, its asymptotic null distribution is not chi-square.

As a simple example, consider the test that the observations come from a single normal distribution versus the alternative that they come from the mixture of two normal distributions. In this case, the mixture model is

$$f_\theta(x) = \pi_1 f_{1,\theta}(x) + (1-\pi_1) f_{2,\theta}(x).$$

Under the null hypothesis we obtain $\pi_1 = 1$, and this solution is on the boundary of the parameter space. This is precisely the reason why the regularity conditions fail.

In certain special cases the asymptotic distribution of $-2\log\lambda$ has been derived. See McLachlan and Basford (1988, pp. 22–24) for examples. In the general formulation, we assume a nested model (i.e., under H_0, the number of distributions is k, whereas under the alternative H_1, there are $m > k$ distributions.

The bootstrap prescription (assuming n observations in the original sample) is as follows: (1) Compute the maximum likelihood estimates of the parameters under H_0; (2) generate a bootstrap sample of size n from the mixture of k distributions defined by the maximum likelihood estimates in step (1); (3) calculate $-2\log\lambda$ for the bootstrap sample; and (4) repeat steps (2) and (3) many times.

The bootstrap distribution of $-2\log\lambda$ is then used to approximate its distribution under H_0. Since critical values are required for the test, estimation of the tails of the distribution are required. McLachlan and Basford recommend at least 350 bootstrap samples.

In order to get an approximate α level test, suppose we have m bootstrap replications with

$$\alpha = 1 - j/(m+1);$$

then we reject H_0 if $-2\log\lambda$ for the original data exceeds the jth smallest value from the m bootstrap replications.

For the simple case of choosing between one normal distribution and the mixture of two normals, McLachlan (1987) performed simulations to show the improvement in the power of the test as m increased from 19 to 99. A number of applications of the above test to clustering problems can be found in McLachlan and Basford (1988).

8.4. CENSORED DATA

The bootstrap has also been applied to some problems involving censored data. Efron (1981a) considers the determination of standard error and confidence interval estimates for parameters of an unknown distribution when the

data are subject to right censoring. He applies the bootstrap to the Channing House data first analyzed by Hyde (1980). Additional discussion of the problem can be found in Efron and Tibshirani (1986).

The Channing House in Palo Alto is a retirement center. The data consist of 97 men who lived in the Channing House during the period from 1964 (when it opened) until July 1, 1975 (when the data were collected). During this period, 46 residents died while in Channing House while the remaining 51 were still alive at the censoring time, July 1, 1975.

The Kaplan–Meier survival curve [see Kaplan and Meier (1958)] is the standard method for estimating the survival distribution, and Greenwood's formula is a common approach to estimating the standard error of the survival curve.

Efron (1981a) compared a bootstrap estimate of the standard error of the Kaplan–Meier curve with Greenwood's formula when applied to the Channing House data. He found very close agreement between the two methods. He also considered bootstrap estimates of the median survival time and percentile method confidence intervals for the median survival time.

Censored data such as the Channing House data consist of a bivariate vector (x_i, d_i), where x_i is the age of the patient at the time of death or at the censoring date and d_i is an indicator variable that is 0 if x_i is censored and 1 if it is not. In the case of the Channing House data, x_i is recorded in months. So, for example, the pair (790, 1) would represent a man who died in the house at age 790 months, while (820, 0) would represent a man living on July 1, 1975, who was 820 months old at that time.

A statistical model for the survival time and the censoring time as described on pp. 64–65 of Efron and Tibshirani (1986) is used to describe the mechanism for generating the bootstrap replications. A somewhat simpler approach treats the pair (x_i, d_i) as an observation from a bivariate distribution F. Simple sampling with replacement from F leads to the same results as the more complex approach which takes account of the censoring mechanism. The conclusion is that the bootstrap provides appropriate estimates of standard errors even when the usual assumptions about the censoring mechanism fail to hold. This result is very reminiscent of the result for bootstrapping paired data in simple regression as compared with bootstrapping residuals from the model.

8.5. *p*-VALUE ADJUSTMENT

In various studies where multiple testing is used or in clinical trials where interim analyses are conducted, the p-value based on one hypothesis test is no longer appropriate. The simultaneous inference approach can be used and is covered well in Miller (1981b). Bounds on the p-value such as the Bonferroni inequality can be very conservative.

A bootstrap approach, as well as a permutation approach to p-value adjustment in the multiple testing framework, was devised by Westfall (1985) in

dealing with multiple binary tests where the binomial model is appropriate and in full generality with examples in Westfall and Young (1993). Software implementation of the method first appeared as PROC MULTTEST in SAS version 6.12. The procedure has been maintained in subsequent updates of SAS including the current version 9.1.

8.5.1. Description of Westfall–Young Approach

As defined in Section 1.3 of Westfall and Young (1993), when we require a family-wise significance level (FWE) to be α, then for each i the adjusted p-value is defined as follows:

$$p_i^a = \inf\{\alpha | H_i \text{ rejected at the FWE} = \alpha\}.$$

This means that p_i^a is the smallest significance level for which one still rejects H_i, given a particular simultaneous test procedure. The concept that is the basis for applying resampling methods in multiple testing situations is the single-step adjusted p-value. Suppose we have k hypothesis tests that are possibly correlated. Let $P_1, P_2, \ldots, P_k$, be the k p-values that have some unknown joint distribution. Then Westfall and Young define the single-step adjusted p-value as follows:

$$p_{is}^a = \mathrm{P}(\min P_j \le p_i | H_o^c).$$

In words: the single-step p-value is the smallest p-value such that at least one of the p-values is no greater than p_i when the alternative hypothesis is true. Westfall and Young (1993) define several types of adjusted p-values, and for each one, they define a bootstrap method for estimating the p-value. For example, Algorithm 2.5 on page 47 of Westfall and Young (1993) provides a method for computing a bootstrap estimate of p_{is}^a. The algorithm works as follows:

Step 0: Set all counters to zero ($i = 1, 2, \ldots, k$).

Step 1: Generate a p-value vector $\mathbf{p}^* = (\mathbf{p}_1^*, \mathbf{p}_2^*, \mathbf{p}_3^*, \ldots, \mathbf{p}_k^*)$ from the same distribution as the original p-values under the full null hypothesis.

Step 2: If min $p_j^* \le p_i$, then add 1 to counter i for each i.

Step 3: Repeat steps 1 and 2 N times.

Step 4: Estimate p_{is}^a as $p_{is}^a(N) = (\text{counter } i)/N$.

8.5.2. Passive Plus DX Example

The PassivePlus DX postmarket approval (PMA) clinical report provides an example where such p-value adjustment using the tools provided by Westfall and Young could have been used. In this and other studies, interim analyses

are performed, usually at a prespecified point in the trial. These days, group sequential designs as well as adaptive designs can formally handle the issue of an appropriate final *p*-value, and some designs are chosen to ensure that a given significance level is maintained and there is no need for a *p*-value adjustment at the end of the trial.

In the Passive Plus example the interim analysis is done mainly for safety. Also, data safety and monitoring boards (DSMB) want to take an early look at the data to see if the benefits outweigh the risks. If the risks are perceived to be too great, the DSMB may require the study to be stopped.

The Passive Plus study is very similar to the Tendril DX study that we described in Section 3.3.1. Passive Plus DX is also a steroid eluting lead. The only difference is the way the leads attach themselves in the heart. Tendril is active fixation; this means that when implanted into the heart, it is screwed into the wall of the chamber.

On the other hand, Passive Plus is a passive fixation lead that is placed by the physician touching but not penetrating the wall. Eventually, fibrous material grows around the lead and keeps it from getting dislodged. Because the Tendril DX leads are active fixations, they are less likely to dislodge, but dislodgement is always a possibility.

Just as with the Tendril leads, there is an approved non-steroid version of the Passive Plus lead which serves as a concurrent active control. It is also a randomized control clinical trial using a 3:1 randomization ratio for treatment (Passive Plus DX lead) to the control (Passive Plus non-steroid lead).

Another aspect of these trials (both the Tendril DX and the Passive Plus DX) is the use of an interim analysis. It is common to have meetings in the early phase of the study to make sure that there are no unusual adverse events that would raise safety issues that could lead to the termination of the study. Often there are data and safety monitoring boards and institutional review boards that want to review the data to get an idea of the risk–benefit tradeoffs in the study.

Commonly, since these are multicenter studies, the investigators want the meetings because they are curious to see how the study is progressing and they want to share their experiences with the investigators. These meetings are also used to ensure that all the centers are complying properly with the protocol. The FDA is also interested in these interim reports, which are generated at the time of these meetings.

The FDA is primarily interested in the safety aspects (report on complications). The study would not be terminated early for better-than-expected performance based on interim report results since the design was a prospective fixed sample size design.

For a group sequential design, the interim points could be chosen prospectively at patient enrollment times when sufficiently good performance would dictate successful termination of the study. However, before the year 2000, the medical device center at the FDA usually doesn't allow such group sequential trials.

In the Passive Plus study, there were two or three interim reports and a final report. Because repeated comparisons were made on capture thresholds for the treatment and control groups, the FDA views the study as a repeated significance test and requires that the sponsor provide a p-value adjustment to account for the multiplicity of tests.

Although the Bonferroni inequality provides a legitimate conservative bound on the p-value, it would be advantageous to the sponsor, St. Jude Medical, to provide an acceptable but more accurate adjustment to the individual p-values. The bootstrap adjustment using PROC MULTTEST in SAS was planned as the method of adjustment.

The results of the test showed that the individual p-values for the main efficacy variables were less than 0.0001 and hence the question became moot. Even the Bonferroni Bound based on three interim reports and one final report would only multiply the p-value by a factor of four, and hence the bound on the adjusted p-value is less than 0.0004. The bootstrap adjusted p-value would be less than the Bonferroni bound.

8.5.3. Consulting Example

In this example a company conducted a clinical trial for a medical treatment in one country, but due to slow enrollment they chose to extend the trial into several other countries. In the first country, which we will call country E, the new treatment appeared to be more effective than the control. But this was not the case in the other countries, which we shall call countries A, B, C, and D.

Fisher's exact test was used to compare the failure rates in each of the other countries with country E. However, this required four pairwise tests that were not part of the original plan, and hence p-value adjustment is appropriate. In this case the Bonferroni bound is too conservative, and so a bootstrap adjustment to the p-value was used. We provide a comparison of the p-values based on a single test referred to as the raw p-value with both the Bonferroni and the bootstrap p-value adjustments. Table 8.1 provides the treatment data for each pair of countries A versus E, B versus E, C versus E, and D versus E. The p-values and the adjusted p-values were determined using PROC

Table 8.1 Comparison of Treatment Failure Rates

Country	Treatment Failure Rate
A	40% (18/45)
B	41% (58/143)
C	29% (20/70)
D	29% (51/177)
E	22% (26/116)

Table 8.2 Comparison of *p*-Value Adjustment

Countries	Raw *p*-Value	Bonferroni *p*-value	Bootstrap *p*-Value
E versus A	0.0307	0.1229	0.0855
E versus B	0.0021	0.0085	0.0062
E versus C	0.3826	1.0000	0.7654
E versus D	0.2776	1.0000	0.6193

MULTTEST in SAS. The results comparing the *p*-value are obtained using the default number of bootstrap replications, which is 20,000.

Table 8.2 shows the conservativeness of the Bonferroni bound. Clearly in each case *p*-value adjustment makes a difference. In the case of country E versus country B the *p*-value is clearly significantly small by either method. In E versus A, however, the raw *p*-value suggested statistical significance at the 5% and 10% significance levels while Bonferroni does not. Note that the bootstrap *p*-value is statistically significant at the 10% level.

These results show that country E is different from at least one and possibly two of the other countries. This could be used by the sponsor to argue the merits of the treatment compared to the control in country E (ignoring the results from the other countries).

8.6. BIOEQUIVALENCE APPLICATIONS

8.6.1. Individual Bioequivalence

Individual bioequivalence is a level of performance expected of a new formulation of a drug to replace an existing approved formulation. Also, this performance measure may be needed for a drug company to market a generic version of a competitor's approved drug. In clinical trials, individual bioequivalence is difficult to access and there are very few statistical techniques to address this issue. In an important paper, Shao, Kübler, and Pigeot (2000) develop a bootstrap procedure to test for individual bioequivalence, and they are able to show mathematically that the bootstrap estimate they generate is consistent.

Pigeot (2001) is a nice summary article that points out the pros and cons of using the jackknife and the bootstrap in biomedical research. She provides the two examples, one to illustrate the value of jackknife procedures and the other to illustrate the virtues of the bootstrap. In this article, Pigeot cautions against the blind application of these techniques because some naive approaches actually fail to work (i.e., fail to be consistent).

The first example is the estimation of a common odds ratio in stratified contingency table analysis. In that example she provides two jackknife estimators of the ratio that are remarkably good at reducing the bias of a ratio estimate.

In the second example she demonstrates how the bootstrap approach of Shao, Kübler, and Pigeot (2000) is successfully applied to demonstrate individual bioequivalence. The bootstrap has been so successful for this problem that it has been recommended in an FDA guidance document on bioequivalence [see FDA Guidance (1997)]. We will now look at Pigeot's bioequivalence example in a little more detail.

In the example, two formulations of a drug are tested for individual bioequivalence using consistent bootstrap confidence intervals. Nothing more sophisticated than Efron's percentile method was used for the confidence intervals, although Pigeot does point out that more sophisticated confidence interval procedures could be used and she refers the reader to Efron and Tibshirani (1993, Chapters 12–14 and 22), Chernick (1999b, Chapter 3), and the article by Carpenter and Bithell (2000) for more details on bootstrap confidence intervals.

Demonstrating bioequivalence amounts to exhibiting equivalent bioavailability as determined by the log transformation of the area under concentration (AUC) versus time curve. The difference of between this measure for the old and new formulations is used to evaluate whether or not the two formulations are equivalent. Calculating this measure for average bioequivalence is fairly easy, but the FDA wants to establish individual bioequivalence and that requires comparing probability estimates rather than group averages. The design favored by the FDA for such studies is what is called the two-by-three crossover design. In this design the old treatment, which is referred to as the reference formulation, is denoted by the letter R. The new formulation is considered the treatment under study and is denoted by T. The plan is for each subject in the study to receive the reference formulation, twice and the treatment once. The only part that involves randomization is the order (sequence) of treatment.

In this case there are three possible sequences (1) RTR, (2) TRR, and (3) RRT. The design is called two by three because the subjects are only randomized to two of the three possible sequences, namely RTR and TRR. The following linear model is assumed for the pharmacokinetic response $Y_{ijk} = \mu + F_l + P_j + Q_k + W_{ljk} + S_{ikl} + \varepsilon_{ijk}$, where μ is the overall mean; P_j is the fixed effect of the jth period with $\Sigma P_j = 0$; Q_k is the fixed effect of kth sequence with $\Sigma Q_k = 0$; F_l is the fixed effect of the lth drug (l is either T or R in our example) and $F_T + F_R = 0$; W_{ljk} is the fixed interaction effect between treatment (drug) sequence and period; S_{ikl} is the random effect of the ith subject in the kth sequence under the lth treatment and the ε_{ijk} are independent identically distributed errors that are independent of the fixed and random effects. Under this model individual bioequivalence is assessed by testing

H_0: $\Delta_{PB} \le \Delta$ versus H_1: $\Delta_{PB} > \Delta$, where Δ_{PB} is defined by $\Delta_{PB} = P_{TR} - P_{RR}$ where $P_{TR} = \text{prob}\,(|Y_T - Y_R| \le r)$ and $P_{r|RR} = \text{prob}(|Y_R - Y_R'| \le r)$ where Δ and r are fixed in advance and Y_R' is the observed response on the second time the reference treatment is given.

Pigeot points out that the FDA guideline, FDA (1997), makes the general recommendation that the hypothesis test is best carried out by constructing an appropriate confidence interval whose lower confidence bound must be greater than Δ to claim bioequivalence. In the bioequivalence framework, it is two one-sided tests each at significance level α that leads to a combined α level test that corresponds to a particular $100(1 - 2\alpha)\%$ confidence interval. In a parametric approach the two one-sided tests are *t* tests.

Pigeot then presents a bootstrap procedure that was recommended in Schall and Luus (1993). Although the method is straightforward, it is inconsistent. See Pigeot (2001) for a detailed description of this bootstrap algorithm. She then points out a very minor modification that makes the confidence interval a bootstrap percentile method interval estimate. She then applies the results from Shao, Kübler, and Pigeot (2000) to claim consistency for the modified result. It points out the care that needs to be taken to make sure that the bootstrap method is correct. The FDA changed its guidance on bioequivalence in 2003, reverting back to using average bioequivalence alone as the attribute for analysis [see Patterson and Jones (2006) for details]. I do not think the issue is permanently settled

8.6.2. Population Bioequivalence

Czado and Munk (2001) present simulation studies comparing the various bootstrap confidence intervals to determine population bioequivalence. The authors point out that aside from draft guidelines, the current guidelines in place for bioequivalence, address average bioequivalence even though the statistical literature has pointed to other measures including individual and population bioequivalence as more relevant when a generic is being considered to replace an approved drug. In an influential paper, Hauck and Anderson (1991) argued that prescribability of a generic drug requires consideration of population bioequivalence.

Population bioequivalence essentially means the population probability distributions are essentially the same. In an unpublished paper, Munk and Czado (1999) developed a bootstrap approach for population bioequivalence. In Czado and Munk (2001), they consider extensive simulations to demonstrate the small sample behavior of various bootstrap confidence intervals and in particular comparing the percentile method bootstrap PC versus the BCa bootstrap. The context of the problem is a 2×2 crossover design. The authors focused on answering the following five questions:

1. How does the PC and BCa intervals compare in small samples with regard to maintaining the significance level and power based on results in small samples.
2. What is the effect on performance of PC and BCa intervals if the populations are normal? When they are very non-normal populations?
3. What is the gain in power when *a priori* period effects are excluded?

4. Does the degree of dependence in the crossover trial influence the performance of the proposed of the proposed bootstrap procedure?
5. How do these bootstrap tests perform compared to the test when the limiting variance is estimated?

The conclusions are as follows: Only bootstrap estimates that are available without a need to estimate the limiting variance were considered. Thus percentile *t* bootstrap procedures were precluded. On the other hand, the PC and the BCa estimates do not require any estimates. Their overall conclusion is that BCa intervals are superior to the PC method, which is very liberal. Gains in power occur when periodic effects can be excluded and when positive correlation within the sequences can be assumed.

8.7. PROCESS CAPABILITY INDICES

In many manufacturing companies, process capability indices that measure how well the production process behaves relative to specifications are popular performance measures. These indices were used by the Japanese in the 1980s as part of their quality improvement movement. In following the Japanese lead based on the teachings of Deming, Juran, Taguchi, and others, the first American companies to apply these methods in the 1990s were the automobile manufacturers. They were motivates by the dramatic loss of business to Japanese car manufacturers. These techniques are now commonly used again in many industries. In fact for medical devices that are regulated by the US Food and Drug Administration (FDA), these methods are now incorporated into the company's product performance qualification process.

In the late 1980s, the 1990s, and beyond, improvement in quality became the goal of many industries in the United States as well as in other countries. The teaching of Deming, Juran, and Taguchi that had previously been ignored were now being followed, and many common quality control techniques including control charts, design of experiments, tolerance intervals, response surface methods, evolutionary operation, and process capability indices are introduced into the manufacturing process at a large number of companies.

Although it has been primarily a positive change, the sweeping movement has unfortunately led to some misuses under the guise of the common popular buzz words including total quality management (TQM) and six sigma. Some companies have been very successful in their efforts to incorporate the six sigma approach, most notably Motorola, AT&T, and General Electric. However, for each success story there has been a well-established statistical research group within the company. There have also been a number of failures where the techniques were implemented too quickly and with lack of understanding and training.

The problem is that a number of these companies would institute the techniques without having or developing the appropriate infrastructure that would

include some statistical expertise. Some procedures that require Gaussian assumptions are blindly applied even in cases where the Gaussian assumptions are clearly not applicable. We shall see how this has led to inappropriate interpretation of many of the process capability indices.

When I was the senior statistician at Pacesetter, there were numerous occasions when engineers would come to me and complain that they could not get the leads, which they were testing to pass the qualification test because either the process capability index was too low or a statistical tolerance interval had at least one endpoint outside of the specification limit, In the case of the tolerance interval it happened in spite of the fact that in 20 to 30 tests, none of the values fell outside the specification limits. This ran counter to their intuition.

Without exception, I would find that the problem was with the statistical method being employed rather than a real qualification problem with the data. In many cases Gaussian tolerance intervals were being used when the data were clearly not Gaussian. So the method did not fit the data; but since the engineers did not know the theory they had made it a standard policy to use the Gaussian tolerance limits not knowing or understanding the statistical limitations to this approach. I found that the distributions in the tolerance interval applications were either highly skewed or very short-tailed. Both characteristics indicate significant departures for the Gaussian distribution and that the use of Gaussian tolerance intervals is a misapplication. A more appropriate approach would be to apply nonparametric tolerance intervals with a larger sample size. I found that this was often the appropriate remedy and I worked to get the SOP changed.

When process capability parameters are computed, numbers like 1.0 or 1.33 have often been taken as standards. But these numbers are justified only based on the Gaussian assumptions that are easily misinterpreted when the data are not Gaussian.

In this case the bootstrap can play a very helpful role by removing the requirement of the normal distributions. Kotz and Johnson (1993) were motivated to write their book on process capability indices to clear up this confusion and to provide a variety of alternative methods in the non-Gaussian case. They also wanted to show that when used appropriately, these methods can be very effectively used in quality assurance and process monitoring. Also more than one simple index is usually needed to properly characterize a complex manufacturing process.

A good historical account can be found in Ryan (1989, Chapter 7). Generally, capability indices are estimated from data and the distributional theory of these estimates is covered in Kotz and Johnson (1993). One of the most popular indices for the case where both upper and lower specification limits exist is C_{pk}.

Let μ be the process mean and let σ be the process standard deviation (the process is assumed to be stable and under control); let LSL denote the lower specification limit and let USL be the upper specification limit. Then C_{pk} is defined as the minimum of (USL-μ)/(3σ) and (m-LSL)/(3σ) and is called a

process capability index. In practice these two limits are estimated by substituting estimates for μ and σ.

For Gaussian distributions, confidence intervals generated and hypothesis tests performed are based on tabulated values for the estimated indices. In practice, the process may have a highly skewed distribution or have at least one short tail.

Non-Gaussian processes have been treated in a variety of ways. Many of these are discussed in Kotz and Johnson (1993, pp. 135–161). The seminal paper of Kane (1986) devotes only a short paragraph to this topic, in which Kane concedes that nonnormality is a common occurrence in practice and that resulting confidence intervals could be sensitive to departures from normality.

Kane states "Alas it is possible to estimate the percentage of parts outside the specification limits, either directly or with a fitted distribution. This percentage can be related to an equivalent capability for a normal distribution."

The point Kane is making is that in some case a lower capability index for a particular distribution (say, one with short tails) has significantly less probability outside the specification limits than for a process with a normal distribution with the same mean and variance. Either the associated probability outside the limit should be specified or the process capability index should be reinterpreted. Kane suggests transforming the index for the actual distribution to the comparable one for a Gaussian distribution. This approach is suggested because managers are now accustomed to thinking of these indices in terms of these standard values for Gaussian indices.

Another approach is taken by Gunter in a series of four papers in the journal *Quality Progress.* In two of Gunter's papers (Gunter, 1989a,b) he emphasizes the difference between "perfect" (precisely normal) and "occasionally erratic" processes (i.e., a mixture of two normal distributions with mixing proportions p and $1 - p$). We take p close to 1 and then the erratic part is due to the occasional sample from the second distribution. The mixture distribution has a different mean and larger variance than the basic distribution.

Gunter considers three types of distributions: (1) a highly skewed distribution, a central chi square with 4.5 degrees of freedom, (2) a heavy-tailed symmetric distribution, a central t distribution with 8 degrees of freedom, and (3) a uniform distribution. Using these exact distributions, Gunter shows what the expected number of nonconforming parts are (i.e., the number of cases outside the 3σ limits of the mean). This is expressed in terms of number of cases per million parts. The distributions are standardized by shift and scale transformations, so that they all have mean 0 and variance 1.

The results are strikingly different, indicating how important the tail of the distribution is to the inference and meaning of the index. For the highly skewed chi square 14,000 are outside the limit but all of them are above the 3σ limit and none are below the -3σ limit. For the uniform distribution there

are no cases outside the limits! We contrast these results to the 2700 cases we would have for a normal distribution with 1350 cases below -3σ and 1350 above 3σ. The t distribution has a significantly larger number of cases in each tail (2000 above 3σ and 2000 below -3σ).

So Gunter's approach shows us that there is a great difference in the probability of falling within the limit, depending greatly on the shape of the distribution. To consider the practical situation of where the true parameters (mean μ and variance σ^2) are unknown and hence estimated from the sample requires that we do simulations. English and Taylor (1990) did extensive simulations using normal, triangular and uniform distributions. We noticed a major difference as can be seen if you consult Table 4.1 of Kotz and Johnson (1993, p. 139). The table in Kotz and Johnson (1993) was taken from English and Taylor (1990). It shows results very similar to what we saw from Gunter's examples.

This points us to one of the factors that caused problems in the use of capability indices. Companies accepted that their indices were characterized by Gaussian data. This was the underlying assumption that they probably didn't even know they were making.

The assumption is implicit when setting a value of 1.33 as a goal for an index. The real goal should be to ensure that only a very small percentage of cases fall outside the specification limits. But instead of trying to directly estimate whether or not we achieve this kind of goal, we estimate C_{pk} from the data and compare it to the values we would obtain from normal distributions.

Another problem with the naive use of sample estimates of the process capability indices is that the companies treat the estimate whose magnitude depends on the number of tests. This uncertainty, which is a function of the sample size, is generally ignored but should not be.

Taking a theoretical approach, Kocherlakota, Kocherlakota, and Kirmani (1992) derive the exact distribution for C_p (a slightly different process capability index) in two specific cases. In one of the cases they look at a mixture of two normal distributions with the same variance. Price and Price (1992) study the expected value of the estimated C_{pk} via simulation for a large number of distributions.

Another approach for dealing with indices from production processes with non-Gaussian distributions is to use the bootstrap. Bootstrapping C_{pk} is fairly straightforward. The limits LSL and USL are fixed. Ordinary bootstrap samples of size n are generated with the sample estimates of the mean and standard deviation calculated for each bootstrap sample.

The index C_{pk}, for example, is then computed for each bootstrap sample based on the above definition with the sample estimates replacing the population parameters. Bootstrap confidence intervals or hypothesis tests for C_{pk} can then be generated usinmethods discussed in Chapter 3. This approach is very general and can be applied to any of the commonly used capability indices and not just for C_{pk}.

Various approaches are discussed in Kotz and Johnson (1993, pp. 161–164) and in the original papers by Franklin and Wasserman (1991) and Price and Price (1992). They involve the use of confidence interval methods that were covered in Chapter 4, including Efron's percentile method and the bias corrected percentile method.

In reviewing this work, Kotz and Johnson (1993) point to Schenker (1985) as an indication of problems with the bias corrected BC method. Apparently no one had taken the step to determine an acceleration constant and produce a BC_a interval or apply bootstrap iteration although Kotz and Johnson (1993) point to Hall's work on bootstrap iteration as an approach that could be used to improve on any of the bootstrap confidence intervals that they did try.

Improvements in statistical software packages has now made it easier to routinely apply process capability indices. As an example, SAS introduced a procedure called PROC CAPABILITY which can be used to calculate various capability indices, provide standard univariate analyses, and test normality assumptions.

The following example illustrates the use of bootstrap confidence intervals in capability studies based on the depth of lesions in catheter ablation procedures. It is clear from the Shapiro–Wilk tests, the stem-and-leaf graph, the boxplots, and the normal probability plots that data depart significantly from the normal distribution to warrant the use of a robust or nonparametric approach to capability index estimation such as the bootstrap.

Figure 8.2 (SAS output from PROC UNIVARIATE) shows the results of the Shapiro–Wilk test and provides the stem-and-leaf graph, boxplots and the normal probability plot for one particular catheter that we identify as catheter number 12. The p-value from the Shapiro–Wilk test is 0.0066. The stem-and-leaf plot and the boxplot show a positive skewness (the sample estimate of skewness is 0.7031).

The data are a summary of 30 samples of lesions generated with catheter number 12 on a beef heart. These catheters are used to create lesions generated from tissue heating caused by RF energy applied to particular locations in a live human heart. The purpose of the lesion is to destroy nerve tissue that would otherwise stimulate the heart, causing improper fast beating in either the atrial or ventricular chamber. If the generated lesion is too small (particularly if it is not deep enough) the nerve tissue will not be destroyed. If it is too deep, there is danger of complications and even a perforation of the heart.

To be sure that the catheter along with the RF generator will produce safe lesions in humans, tests are performed on live animals or on beef hearts from slain cattle. There is no way to determine the width, depth, and length of the lesion inside a living heart. Consequently, a surrogate measure of effectiveness of the catheter and the RF generator is obtained through the measurements of the lesions generated in a beef heart.

In these tests the length, width, and depth of each lesion are measured. The performance of the catheter and the generator in these tests is a basis for predicting their performance for the catheter in humans. For catheter 12, the

Univariate Procedure Summary for Navistar DS Lesion Dimensions

08:06 Wedensday, November 25, 1998

Univariate Procedure

Variable = DEPTH

Moments				Quantiles (Def = 5)				Extremes			
N	30	Sum Wgts	30	100% Max	9	99%	9	Lowest	Obs	Highest	Obs
Mean	6.65	Sum	199.5	75% Q3	7	95%	8	5(	1)	7(	30)
Std Dev	0.852481	Variance	0.726724	50% Med	6.75	90%	8	5.5(	5)	8(	8)
Skewness	0.703112	Kurtosis	0.819517	25% Q1	6	10%	6	6(	26)	8(	9)
USS	1347.75	CSS	21.075	0% Min	5	5%	5.5	6(	24)	8(	13)
CV	12.81927	Std Mean	0.155641			1%	5	6(	23)	9(	14)
T:Mean = 0	42.72652	Pr > |T|	0.0001	Range	4						
Num ^= 0	30	Num > 0	30	Q3-Q1	1						
M(Sign)	15	Pr >= |M|	0.0001	Mode	6						
Sgn Rank	232.5	Pr >= |S|	0.0001								
W: Normal	0.895608	Pr < W	0.0066								

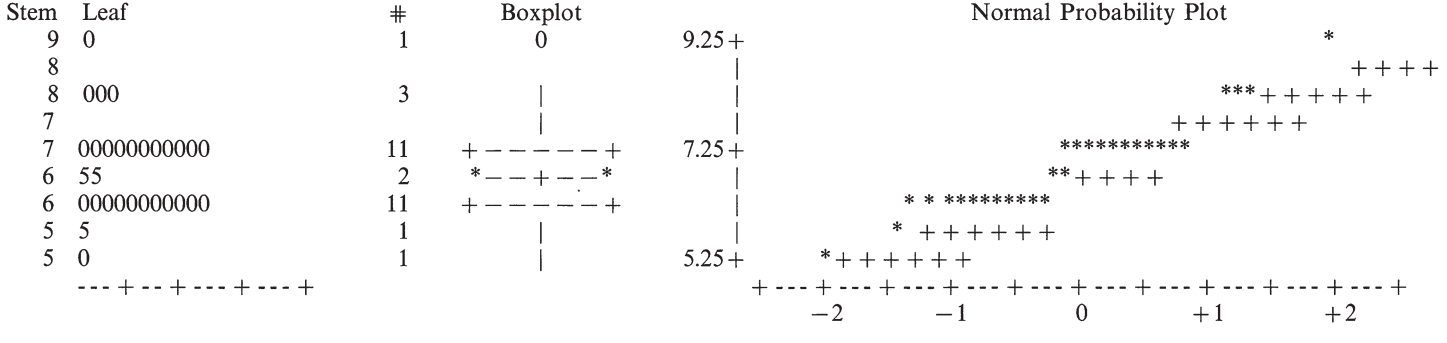


Figure 8.2 Summary statistics for Navistar DS lesion depth dimension from original SAS univariate procedure output.

average lesionb depth was 6.650 mm with a standard deviation of 0.852 mm. The minimum value of the lesion depth was 5.000 mm and the maximum value was 9.000 mm. Similar results were found for other catheters.

In theory, there should be a maximum value beyond which lesion size would be so large that the patient would be in danger of complications such as perforation. On the other hand, if the lesion is too small in depth, it would surely be to be ineffective. Consequently upper and lower specification limits could possibly be determined. In practice, no specific limits have been determined and the correlation with lesions generated internally in humans remains unclear.

For illustrative purposes, we shall assume an upper specification limit of 10 mm and a lower limit of 4 mm with the target value at 7 mm. Figure 8.3 is the SAS output from PROC CAPABILITY, which shows the estimate of the C_{pk} index and other capability indices based on the 30 sample lesions in a beef heart using catheter 12. For C_{pk}, we observe an estimate of 1.036. We shall use the bootstrap to estimate the variability of this estimate and to provide confidence intervals.

Next, we generate bootstrap samples of size 30 to determine the sampling distribution of the capability index. The bootstrap percentile method confidence interval is then generated to illustrate the technique.

The C_{pk} index is a natural index to use here since the specification limits are two-sided. We want a strong enough lesion to be sure that it is effective. Yet we do not want the lesion to be too deep as to make it cause a perforation or other complication.

Another problem in the medical device industry that could involve a two-sided specification is the seal strength of sterile packaging. The seal requires enough strength so that the sterility of the package will not be compromised in shipping. On the other hand, it must not be too well sealed to make it difficult to open when it is time to use it.

We now return to the lesion data example. We apply bootstrap resampling and Efron's percentile method approach. A bootstrap histogram for C_{pk} is shown in Figure 8.4. I used the statistical package Resampling Stats, a product of Resampling Stats, Inc., to generate the 95% bootstrap percentile method confidence limits for C_{pk}. The computation of the capability indices required generating some code in the Resampling Stats Programming Language. The actual code that I used is presented as Figure 8.5 for the percentile method.

The figures show both the code used in the execution of the program and the output of the run. In each case I used 10,000 bootstrap replications to produce the confidence limits and the program executed very rapidly (19.7 seconds total execution time for the percentile method). Resampling Stats provides a very readable user's manual [see Resampling Stats (1997)].

The results showed that the estimated C_{pk} was 1.0362. The percentile method confidence interval is [0.84084, 1.4608], showing a great deal of uncertainty in the estimate. Perhaps to get a good estimate of C_{pk} a sample size much larger than 30 should be used. This example and other issues with C_{pk}

Capability Indices for Lesion Data

08:06 Wedensday, November 25, 1998

CAPABILITY

Variable = DEPTH

Moments				Quantiles (Def=5)				Extremes			
N	30	Sum Wgts	30	100% Max	9	99%	9	Lowest	Obs	Highest	Obs
Mean	6.65	Sum	199.5	75% Q3	7	95%	8	5(	1)	7(	30)
Std Dev	0.852481	Variance	0.726724	50% Med	6.75	90%	8	5.5(	5)	8(	8)
Skewness	0.703112	Kurtosis	0.819517	25% Q1	6	10%	6	6(	26)	8(	9)
USS	1347.75	CSS	21.075	0% Min	5	5%	5.5	6(	24)	8(	13)
CV	12.81927	Std Mean	0.155641			1%	5	6(	23)	9(	14)
T:Mean=0	42.72652	Pr>|T|	0.0000	Range	4						
Sgn Rank	232.5	Pr>=|S|	0.0000	Q3-Q1	1						
Num ^=0	30			Mode	6						
W:Normal	0.895608	Pr<W	0.0066								

Specifications			
LSL	4	% < LSL	0
Target	7		
USL	10	% > USL	0
		% Between	100

Indices			
CPL	1.036191	CPU	1.309902
CP	1.173046	CPK	1.036191
K	0.116667	CPM	1.085148

WARNING: Normality rejected at $\alpha = 0.05$

Figure 8.3 Capability statistics for Navistar DS lesion depth dimension from original SAS capability procedure output for catheter 12.

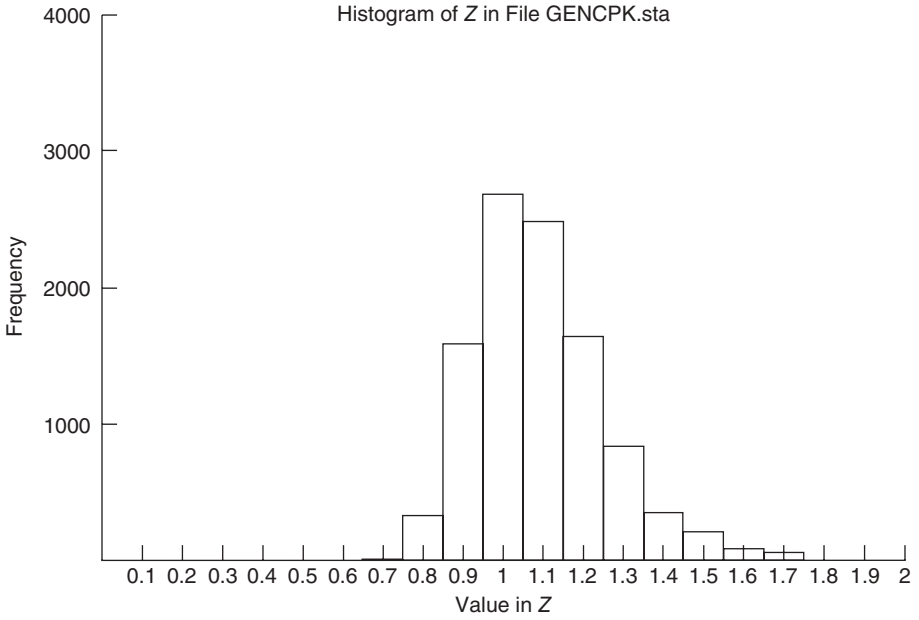


Figure 8.4 **A** bootstrap histogram for Navistar DS lesion depth dimension based on an original sample of 30 for catheter number 12.

is covered in Chernick (1999a). In doing the research for Chernick (1999a), I discovered that an unpublished paper, Heavlin (1988), produces approximate two-sided 95% confidence intervals in the Gaussian case. As we have seen, the Gaussian assumption is not reasonable. So we would expect Heavlin's interval to differ substantially from the bootstrap version. In fact, when I applied Heavlin's method to the same lesion data, I got [0.7084, 1.3649]. Note that in comparison to the bootstrap percentile method, this interval is shifted to the left and is slightly wider.

8.8. MISSING DATA

Many times in clinical studies complete data are not recorded on all patients. This is particularly true when quality-of-life surveys are conducted. In such studies many standard statistical analyses can be conducted only on complete data. Incomplete data can be analyzed directly using the likelihood approach along with the EM algorithm.

Alternatively, imputation of all missing values can lead to an artificial complete data set that can be analyzed by standard methods. However, even if the method of imputation works appropriately as in the case where a model for the missing data is right, it is well known that when only a single imputation

Start execution.

```
DATA (5.0 6.0 6.5 7.0 5.5 6.0 6.5 8.0 8.0 7.0 7.0 6.0 8.0 9.0 6.0 6.0 6.0 6.0 7.0 7.0 6.0 7.0 6.0
6.0 7.0 6.0 7.0 7.0 7.0 7.0) DEPTH
MAXSIZE Z 15000
MEAN DEPTH MD
STDEV DEPTH SIGD
SUBTRACT 10 MD USLNUM
SUBTRACT MD 4 LSLNUM
MULTIPLY SIGD 3 DENOM
DIVIDE USLNUM DENOM CPU
DIVIDE LSLNUM DENOM CPL
PRINT CPU CPL MD SIGD
CONCAT CPU CPL CP
MIN CP CPK
PRINT CPK
REPEAT 10000
  SAMPLE 30 DEPTH BOOTD
  MEAN BOOTD MDS
      STDEV BOOTD SIGD$
  SUBTRACT 10 MD$ USLNUM$
        SUBTRACT MD$ 4 LSLNUM$
      MULTIPLY SIGD$ 3 DENOM$
  DIVIDE USLNUM$ DENOM$ CUP$
        DIVIDE LSLNUM$ DENOM$ CPL$
        CONCAT CPU$ CPL$ CP$
  MIN CP$ CPK$
  SCORE CPK$ Z

END
HISTOGRAM Z
PERCENTILE Z (2.5 97.5) K
PRINT CPK K
CPU  =  1.3099
CPL  =  1.0362
MD   =  6.65
SIGD =  0.85248
CPK  = 1.0362
```

Vector no. 1: z

Bin Center	Freq	Pct	Cum Pct
0.7	10	0.1	0.1
0.6	306	3.1	3.2
0.9	1554	15.5	18.7
1	2668	26.7	45.4
1.1	2461	24.6	70.0
1.2	1603	16.0	86.0
1.3	804	8.0	94.1
1.4	327	3.3	97.3
1.5	184	1.8	99.2
1.6	52	0.5	99.7
1.7	23	0.2	99.9
1.8	5	0.1	100.0
1.9	3	0.0	100.0

Note: Each bin covers all values within 0.05 of its center.

```
CPK  =  1.0362
K    =  0.84084     1.4608

Successful execution (19.7 seconds)
```

Figure 8.5 Resampling Stats code and output for the bootstrap percentile method 95%.

is made for each missing value, it is well known that the method underestimates the variability because the uncertainty caused by imputing values is usually not accounted for. The remedies for this include: (1) use a mixed linear model to account for the variability of the missing data through the random effects, (2) use the model and knowledge of the missing data mechanism to adjust the variance estimate, and (3) apply multiple imputation methods to generate a distribution of values for the missing data so as to properly incorporate the uncertainty due to imputation.

Other approaches involve smoothing (interpolation) and extrapolation to impute values for individual patients. Although these methods have drawbacks, there are situations where they can work. One technique that was very commonly used in the past in the pharmaceutical industry is called last observation carried forward (LOCF). It is still used in some cases. However, under most circumstances the estimates are very biased. If it is known that the variable is not increasing or decreasing very much and dropout for lack of efficacy is not a serious problem, the bias of LOCF may not be very large but the variability will still be underestimated. Nevertheless, there are some rare situations where it may still be acceptable. Also, the FDA may sometimes accept LOCF imputation for the intention to treat (ITT) analysis as long as another analysis is performed using the worst-case treatment for the missing values.

Rubin (1987) is the authoritative text on multiple imputation, and Little and Rubin (1987) is one of the best texts dealing with missing data. In a real clinical study at Eli Lilly, Stacy David compared multiple imputation with other imputation methods using a Bayesian bootstrap. This was first reported at an FDA sponsored workshop in Crystal City, Virginia, in September 1998.

David's study showed the value of multiple imputation for certain missing value problems and particularly showed the weaknesses of LOCF. Around the same time, an Amgen study also reported advantages to multiple imputation in the framework of a realistic clinical trial. The Amgen studies and perhaps some other factors led the FDA to recommend multiple imputation in situations where the missing data have a serious effect on the trial results. Many studies at that time and since have shown the weakness of LOCF. A motivating paper that led to the Amgen study and the subsequent work at Eli Lilly is Lavori et al. (1995). An excellent very recent treatment and discussion of the missing data problems in pharmaceutical clinical trials data is Chapter 12, "Analysis of Incomplete Data" (pp. 313–359), by Molenberghs et al. in *Pharmaceutical Statistics Using SAS: A Practical Guide (2007).*

8.9. POINT PROCESSES

A point process is a collection of events or occurrences over a particular period of time. Examples include the time of failures for a medical device, the time to recurrence of an arrhythmia, the lifetime of a patient with a known

disease, and the time to occurrence of an earthquake. There is a plethora of possible examples.

Point processes are studied either in terms of the time between events or by the number of events occurring in a given interval of time. The simplest point process is the Poisson process, which occurs under simple assumptions including the following: (1) rarity of events and (2) the expected number of events in an interval length of time t equals λt, where λ is known as the rate or intensity of the process.

The simple Poisson process gets its name because it has a Poisson distribution with parameter λ for the number of events in an interval of unit length. For a Poisson process the times between events are independent and identically distributed negative exponential random variables with rate parameter λ.

The mean time between events for a Poisson process is λ^{-1}. The Poisson process is a special case of larger classes of point processes including (1) stationary processes, (2) renewal processes, and (3) regenerative processes.

In addition to being able to define point processes on the time line, we can define point processes on the time line, and we can define them in higher dimensions where time is replaced by spatial coordinates. In general, point processes can be determined by their joint probability distributions over subsets (intervals) of the time line or regions in space.

There are also classes of nonstationary point processes. A point process that has its instantaneous rate that varies with time but still has joint distributions that are Poisson and independent over disjoint regions or intervals is called a nonhomogeneous or inhomogeneous Poisson process. Inhomogeneous Poisson processes are nonstationary.

Detailed theoretical treatment of point processes can be found in Daley and Vere-Jones (1988) and Bremaud (1981). Murthy (1974) presents renewal processes along with medical and engineering applications. Bootstrap applications to homogeneous and inhomogeneous Poisson processes are presented in Davison and Hinkley (1997, pp. 415–426).

There are some very simplistic approaches to resampling a point process that Davison and Hinkley describe. The simplest is to randomly select a sample from the set of observed events. This relies on the independence assumption for Poisson and hence is somewhat restrictive.

A different approach for an inhomogenous Poisson process would be to compute a smoothed estimate of the intensity (rate) function based on the original data and then generate realizations of the inhomogeneous Poisson process with the estimated intensity function via Monte Carlo. This would be a form of parametric bootstrapping. Davison and Hinkley take such an approach with a point process of neurophysiological data.

Figure 8.6 is taken from Davison and Hinkley (1997). The data were generated by Dr. S. J. Boniface of the Clinical Neurophysiology Unit at the Radcliff Infirmary in Oxford, England. It is a study of human subjects responding to a stimulus. The stimulus was applied 100 times and points in the figure indicate

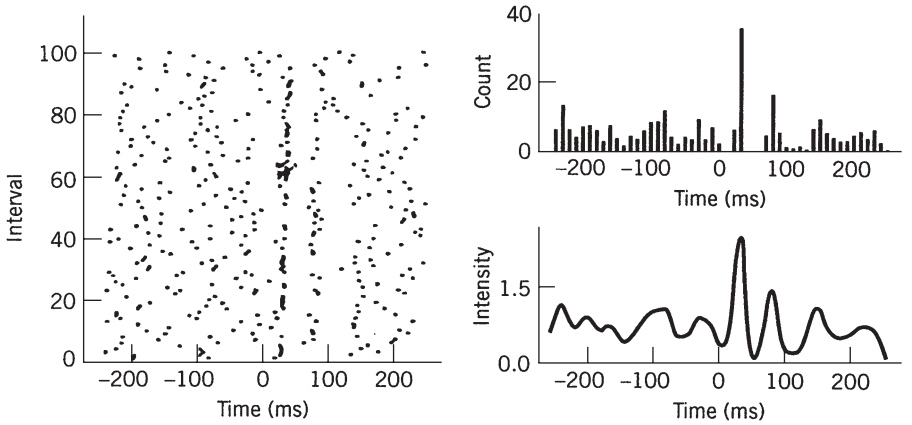


Figure 8.6 Neurophysiological point process. [From Davison and Hinkley (1997, Figure 8.15, p. 418), with permission from Cambridge University Press.]

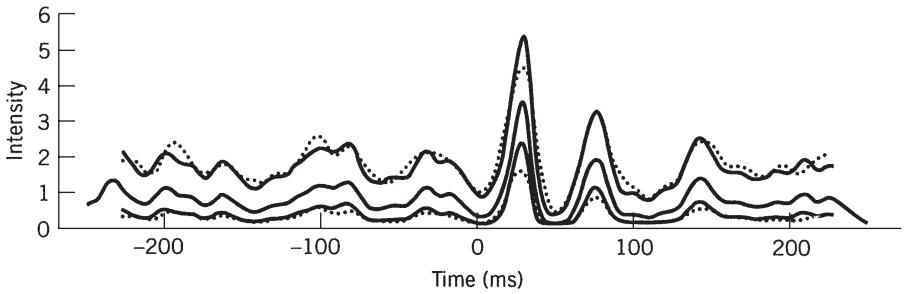


Figure 8.7 Confidence bands for the intensity of neurophysiological point process data. [From Davison and Hinkley (1997, Figure 8.1.6, p. 421), with permission from Cambridge University Press.]

times in a 250-msec interval for which the response (firing of a motoneuron) was observed. Theoretical considerations indicate that the process is obtained by superimposed occurrences (the upper right panel) and a rescaled kernel estimate of the intensity function $\lambda(t)$.

We would like to put confidence bounds on this estimated intensity function. Two-sided 95% bootstrap confidence bands are presented in Figure 8.7.

8.10. LATTICE VARIABLES

Lattice variables are variables that are defined on a discrete set, and so they apply to a lot of discrete distributions including the binomial. The first key

theorems on the consistency of the bootstrap mean by Singh (1981) and Bickel and Freedman (1981) involved requirements that the observations are nonlattice. See Theorem 11.1 of Lahiri (2006, p. 235) for the theorem and a proof. In addition, under the conditions of the theorem, it is known that based on Edgeworth expansions, the second-order correctness of the bootstrap can be demonstrated showing that the bootstrap is superior to the simplest normal approximation. Unfortunately, the nonlattice requirement is necessary.

Theorem 11.2 of Lahiri (2006, p. 237) shows consistency in the lattice case, but unfortunately the rate of convergence is as slow as the normal approximation and therefore the bootstrap has no advantage over the simple normal approximation in the lattice case this applies to the binomial as a special case.

I have not seen bootstrap applications with lattice variables, and it is probably because of the slow rate of convergence.

8.11. HISTORICAL NOTES

The application of the bootstrap to kriging was done by Switzer and Eynon and documented in unpublished work at Stanford [later published as Eynon and Switzer (1983)]. The results were also summarized in Diaconis and Efron (1983), who referred to the unpublished work. Cressie (1991, 1993) deals with kriging and bootstrapping in the case of spatial data and is still one of the best sources for information on resampling with spatial data. The recent text by Lahiri (2003a) covers recent developments in spatial data analysis and the latest developments in resampling applications.

Hall (1985, 1988c) provide thoughtful accounts for applying bootstrap approaches for spatial data. There are many excellent texts that deal with spatial data, including Bartlett (1975), Diggle (1983), Ripley (1981, 1988), Cliff and Ord (1973, 1981), and Upton and Fingleton (1985, 1989). Matheron (1975) provides some fundamental probability theory for spatial data. Mardia, Kent, and Bibby (1979) partially deal with spatial data in the context of multivariate data analysis.

The idea of applying the bootstrap to determine the variability of subset selection procedures in logistic regression (the same idea can also be applied to multiple regression) was due to Gong. Her analysis of Dr. Gregory's procedure was an important part of her Ph.D. dissertation at Stanford. Results on this theme can be found in Efron and Gong (1981, 1983) and Gong (1982, 1986).

Gong (1986) is the essence of her dissertation. As part of their theme to emphasize the variety of applications and the power of computer-intensive methods, Diaconis and Efron (1983) includes some reference to Gong's work.

Miller (1990) deals with model selection procedures in regression. A general approach to model selection based on information theory is called stochastic complexity and is covered in detail in Rissanen (1989).

Applications of the bootstrap to determine the number of distributions in a mixture model can be found in McLachlan and Basford (1988). A simulation study of the power of the bootstrap test for a mixture of two normals vs a single normal distribution is given in McLachlan (1987).

Work on censored data applications of the bootstrap began with Turnbull and Mitchell (1978), who considered complicated censoring mechanisms. Efron (1981a) is a seminal paper on the subject.

The more advanced confidence interval procedures such as Efron (1987) and Hall (1988b) can usually provide improvement over the procedures discussed in Efron (1981a). Reid (1981) provides an approach to estimating the median of a Kaplan–Meier survival curve based on influence functions.

Akritas (1986) compared variance estimates for median survival using the Efron (1981a) approach with the Reid (1981) approach. Altman and Andersen (1989), Chen and George (1985), and Sauerbrei and Schumacher (1992) apply case resampling methods to survival data models such as the Cox proportional hazards model. There is no theory yet to support this approach.

Applications of survival analysis methods and reliability studies can be found in Miller (1981a) and Meeker and Escobar (1998), to name two out of many. Csorgo, Csorgo, and Horvath (1986) apply results for empirical processes to reliability problems.

The bootstrap application to p-value adjustment was first given in Westfall (1985) and was generalized by Westfall and Young (1993). Their approach using both bootstrap and permutation methods have been implemented in statistical software. In particular, PROC MULTTEST has been available in SAS since Version 6.12. This SAS procedure provides bootstrap and permutation resampling approaches to multiple testing problems using p-value adjustment, along with various classical bounds on the family-wise p-value (e.g., Sidak and Bonferroni) as illustrated in Section 8.5 for the Bonferroni. Bootstrap applications to multiple testing problems are also covered in Noreen (1989).

Initial work on the distribution of process capability indices estimates is due to Kane (1986). The bootstrap work was described in Franklin and Wasserman (1991, 1992, 1994), Wasserman, Mohsen, and Franklin (1991), and Price and Price (1992). A review of this work and a general account of capability indices can be found in Kotz and Johnson (1993).

The Bayesian bootstrap was originated by Rubin (1981). Dempster, Laird, and Rubin (1977) is the classic seminal article that introduced the EM algorithm to most statisticians as a way to handle many important missing data problems. They were the ones to popularize the method and provide nice solutions to several important problems. A detailed up-to-date account can be found in McLachlan and Krishnan (1997).

Rubin (1987) is the treatise on multiple imputation as an approach to missing data problems. Efron (1994) is a key reference on resampling approaches to missing data problems. Rubin and Schenker (1998) present an account of imputation methods including the usage of the Bayesian bootstrap.

Key papers include their original works (Rubin and Schenker 1986, 1991). Application of bootstrap techniques to imputation problems in survey sampling is presented in Shao and Sitter (1996).

Press (1989) is a good reference for the Bayesian approach to inference and Gelman, Carlin, Stern, and Rubin (1995) give a modern treatment with the recent advances in computation of Bayesian a posteriori distributions using Markov chain Monte Carlo methods. Maritz and Lwin (1989) present empirical Bayes' methods as does the more recent text, Carlin and Louis (1996).

For inhomogeneous Poisson processes, examples sketched in the chapter along with an outline of the related theory can be found in Cowling, Hall, and Phillips (1996). Ventura, Davison, and Boniface (1997) describe another method with an application to neurophysiological data. Diggle, Lange, and Benes (1991) provide an application of bootstrap to point process problems in neuroanatomy. Point processes are also referred to as counting processes by some authors.

Reliability analysis and life data (survival) analysis involve the study of censored survival times. These can be studied in terms of time to event or the number of events in an interval.

The relationship $P\{N(t) \geq n\} = P\{X_1 \leq t, X_2 \leq t, \ldots, X_n \leq t\}$, where $N(t)$ is the number of events in the interval [0,t] and $X_1, X_2, \ldots, X_n$ are the times for the first n events relates the counting process $N(t)$ to the time to event observations $\{X_i\}$ $i = 1, 2, 3, \ldots, n$. This means that there is a 1–1 correspondence between the counting process and the time-to-event distribution when the events are, for example, IID. The most well-known case is when the $\{X_i\}$ $i = 1, 2, 3, \ldots, n$, are IID negative exponential and then the corresponding point (counting) process is a homogeneous Poisson process. In addition to the 1–1 correspondence, the counting process determines the distribution of the time to event process and vice versa. One book that takes the counting process approach is Anderson, Brogan, Gill, and Keidling (1993).

CHAPTER 9

When Bootstrapping Fails Along with Some Remedies for Failures

The issue that we raise in the title of this chapter shall be addressed mainly in the narrow context of the nonparametric version of the bootstrap (i.e., sampling with replacement from the empirical distribution function F_n). One exception to this is the case of unstable autoregressive processes covered in Section 9.6. There has been some controversy over exactly what situations are appropriate for bootstrapping.

In complex survey sampling problems and in certain regression problems, some researchers have argued against the use of the bootstrap. On the other hand, early articles such as Diaconis and Efron (1983) and Efron and Gong (1983) painted a very rosy picture for the general application of the bootstrap.

It is not my intention to resolve the controversy here. The reader is referred to LePage and Billard (1992) for research that addresses this issue but emphasizes ways to extend the limits of the bootstrap. Mammen (1992b) provides mathematical conditions that allow the bootstrap to work. In the seven years that have past since the publication of the first edition, most of the naive examples presented in this chapter have been remedied by modifications to the bootstrap and so I now include, in addition to the examples of the inconsistency for the simple bootstrap procedures in Sections 9.2–9.6, modifications that make the bootstrap estimates consistent, Needless to say, this issue is still not completely resolved, and hopefully the limitations of the bootstrap will be better defined after more research has been completed.

There are, however, for practitioners and researchers, known results about inconsistency of bootstrap estimates that they should know about. In Sections

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

9.2–9.6 we shall briefly describe examples where the "bootstrap algorithm" fails (i.e., replacing population parameters with sample estimates and replacing sample estimates with estimates from a bootstrap sample). These examples are intended to caution the practitioner against blind application of bootstrap methods.

Before discussing these examples, we will point out in Section 9.1 why it is unwise to apply the bootstrap in very small samples. In general, if the sample size is very small (e.g., less than 10 in the case of a single parameter estimate), it is unwise to draw inferences from the estimate or to rely on an estimate of standard error to describe the variability of the estimate.

Nothing magically changes when one is willing to apply the bootstrap Monte Carlo approximation to the data. Unfortunately, the name "bootstrap" may suggest to some "getting something for nothing." Of course, the bootstrap cannot do that! The proper interpretation should be "getting the most from the little that is available."

9.1. TOO SMALL OF A SAMPLE SIZE

For the nonparametric bootstrap, the bootstrap samples are, of course, drawn from a discrete set. Many exact and approximate results about these "atoms" of the bootstrap distribution can be found in results from classical occupancy or multinomial distribution theory [see Chernick and Murthy (1985) and Hall (1992a, Appendix I)]. In Chapter 2, we saw examples where bootstrap bias correction worked better than cross-validation in the estimation of the error rate of a linear discriminant function. Although we have good reasons not to trust the bootstrap in very small samples and theoretical justification is asymptotic, the results were surprisingly good even for sample sizes as small as 14 in the two-class problem. A main concern in small samples is that with only a few values to select from, the bootstrap sample will underrepresent the true variability since observations are frequently repeated and bootstrap samples, themselves, can repeat.

The bootstrap distribution is defined to be the distribution of all the possible samples of size n drawn with replacement from the original sample of size n. The number of possible samples or atoms is shown in Hall (1992a, Appendix I) to be

$$\binom{2n-1}{n} = (2n-1)!/[n!(n-1)!].$$

Even for n as small as 20, this is a very large number and consequently, when generating the resamples by Monte Carlo, the probability of repeating a particular resample is small. For n = 10, the number of possible atoms is 92,378.

As a concrete example, Hall (1992a, Appendix I) shows that when $n = 20$ and the number of bootstrap repetitions B is 2000, the probability is greater than .95 that none of the bootstrap samples will repeat. This suggests that for many practical problems, the bootstrap distribution may be regarded as though it were continuous. It also may suggest why researchers have found that smoothing the empirical distribution is less help than was originally expected.

Because the bootstrap distribution grows so rapidly with increasing n, exact computations of the bootstrap estimate are usually not possible (there are, of course, some exceptions where theoretical tricks are applied, e.g., the sample median) and the Monte Carlo approximations are necessary. For very small n, say $n \leq 8$, it is feasible to calculate the exact bootstrap estimate [see Fisher and Hall (1990) for a discussion of such algorithms and/or Diaconis and Holmes (1994), who do complete enumeration using Gray codes].

The practitioner should be guided by common statistical practice here. Samples of size less than 10 are usually too small to rely on sample estimates, even in "nice" parametric cases. So we should expect that such sample sizes are also too small for bootstrap estimates to be of much use.

In many practical contexts, the number 30 is used as a "minimum" sample size. Justification for this can be found by noting the closeness of the central t distribution with $n - 1$ degrees of freedom to the standard Gaussian (normal distribution with mean equal to zero and variance equal to one) when $n \geq 30$. This suggests that when sampling from Gaussian populations, the effect of the estimated standard deviation has essentially disappeared. Also the normal approximation to distributions such as the binomial or the sum of independent uniform random variables is usually very accurate for $n \geq 30$. In the case of the binomial, we must exclude the highly skewed cases where the success probability p is close to either 0 or 1.

It would seem that for many practical cases, sample size considerations should not be altered when applying the bootstrap. In nonparametric problems, larger sample sizes are required to make up for the lack of information that is implicit in parametric assumptions. Although it is always dangerous to set "rules of thumb" for sample sizes, I would suggest that in most cases it would be wise to take $n \geq 50$, if possible.

The best rule on the Monte Carlo approximation is to take $B = 100$ (at least) for bias and standard error estimation and take $B = 1000$ for confidence intervals. One can always cite exceptions to these rules in the literature, and variance reduction techniques may help to reduce B as we have discussed in Chapter 7. Also, keep in mind the discussion regarding the results of Booth and Sarkar (1998) and Efron (1987). In light of the speed of computing today and the suggestions of Booth and Sarkar (1998), I would boost my recommendation to $B = 1000$ for standard error estimation and $B = 10{,}000$ for confidence intervals. This was my recommendation in 1999. In 2007 the computers are so much faster that Monte Carlo replications of 100,000 or more are done routinely except in very complex modeling situations.

9.2. DISTRIBUTION WITH INFINITE MOMENTS

9.2.1. Introduction

Singh (1981) and Bickel and Freedman (1981) showed that if $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with finite second moments and if $Y_1, Y_2, \ldots, Y_n$ are chosen by simple random sampling with replacement from the sample $X_1, X_2, \ldots, X_n$ then letting

$$H_n(x, \omega) = P\left(\frac{\bar{Y}_n - \bar{X}_n}{S_n} \le x | X_1, X_2, \ldots, X_n | \right),$$

where

$$\bar{Y}_n = n^{-1} \sum_{i=1}^{n} Y_t, \bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i, S_n^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)$$

and $\Phi(x)$ is the cumulative standard normal distribution, we obtain

$$\sup |H_n(\boldsymbol{x}, \omega) - \Phi(\boldsymbol{x})| \to 0, \qquad -\infty < \boldsymbol{x} < \infty$$

with probability 1 as $n \to \infty$.

Here ω denotes the random outcome that leads to the values $X_1, X_2, \ldots, X_n$ and $H_n(\boldsymbol{x}, \omega)$ is a random probability distribution. For fixed $X_1, X_2, \ldots, X_n$ (i.e., a particular ω), $H_n(\boldsymbol{x}, \omega)$ is a cumulative probability distribution. It is well known (i.e., the central limit theorem) that

$$G_n(x) = P\left(\frac{\bar{X}_n - \mu}{\sigma} \le x \right) \text{converges to } \Phi(\boldsymbol{x}).$$

The bootstrap principle replaces μ and σ with $\bar{X}_n$ and S_n and replaces $\bar{X}_n$ with $\bar{Y}_n$. So we would hope that $H_n(\boldsymbol{x}, \omega)$ would also converge to $\Phi(\boldsymbol{x})$ for almost all ω. That is what the result of Bickel and Freedman (1981) and Singh (1981) tells us.

Athreya (1987b) asks whether the same kind of result holds in cases where the variance is not finite and so the central limit theorem does not apply. In the case when $X_1, X_2, \ldots, X_n$ have a distribution F satisfying

$$1 - F(\boldsymbol{x}) \sim \boldsymbol{x}^{-\alpha} L(\boldsymbol{x})$$

and

$$f(-x) \sim c x^{-\alpha} L(x)$$

as $x \to \infty$ where L is a slowly varying function as $x \to \infty$ and c is a nonnegative constant, $\bar{X}_n$ appropriately normalized converges to a stable law for $0 < \alpha \leq 2$. When $\alpha = 2$, we get the usual central limit theorem. For $\alpha < 2$ the variance does not exist, but the norming constants and the limiting distribution are well-defined. See Feller (1971) for more details.

9.2.2. Example of Inconsistency

Theorem 1 of Athreya (1987b) proves the basic result for $1 < \alpha < 2$, in the case when $0 < \alpha \leq 1$ was given in an earlier unpublished report. The theorem tells us that when we apply the bootstrap algorithm to the appropriately normalized mean, we get convergence to a random probability distribution rather than a fixed probability distribution and so if

$$H_n(x, \omega) = P[T_n \leq x | X_1, X_2, \ldots, X_n]$$

where

$$T_n = n x_{(n)}^{-1} (\bar{Y}_n - \bar{X}_n)$$

and

$$X_{(n)} = \max(X_1, X_2, \ldots, X_n)$$

and let $G(x)$ be the limiting stable law for the mean, we would have hoped that

$$\sup |H_n(x, \omega) - G(x)| \to 0$$

$$-\infty < x < \infty$$

with probability one. Unfortunately, since $H_n(x, \omega)$ converges to a random probability distribution, we do not get the hoped for result. No fixed distribution other than G can work either since the resulting asymptotic distribution is random.

A similar example for the maximum was given by Bickel and Freedman (1981) and Knight (1989). Angus (1989) has generalized their counterexample to the minimum and maximum of independent identically distributed random variables, which we discuss in the next section.

9.2.3. Remedies

Athreya (1987b) points out that difficulties with the bootstrap mean in heavy-tailed situations can be overcome by the use of trimmed means or a smaller number of observations in the resample (less than n but still of order n).

Intuitively, an appropriately trimmed mean could have the same limiting mean value, but the trimming could lead to a lighter-tailed sampling distribution and so the conditions for the mean with finite variance would apply. The *m*-out-of-*n* bootstrap is a simple way to bootstrap that amazingly solves the consistency problem in a number of cases. Now as long as we choose $m = o(n)$ (i.e., $m \to \infty$ as $n \to \infty$ but $m/n \to 0$ as m and $n \to \infty$) the bootstrap mean is consistent. A proof of the convergence of the bootstrap mean when applying an *m*-out-of-*n* bootstrap was given in Athreya (1987b). In two other papers, Gine and Zinn (1989) and Arcones and Gine (1989) develop additional results on the validity of bootstrapping the sample mean, depending on moment conditions and the resample size $m < n$.

9.3. ESTIMATING EXTREME VALUES

9.3.1. Introduction

Suppose we have a sample size of n from a probability distribution F which is in the domain of attraction of one of the three extreme value types. This assumption requires that certain conditions on the tail behavior of F are satisfied. See Galambos (1978) for details on the limiting distributions and Gnedenko's theorem for the maximum or minimum of a sequence of independent identically distributed random variable. By Gnedenko's theorem, the maximum $X_{(n)}$ or the minimum $X_{(1)}$ of the sequence of observations X_1, $X_2, X_3, \ldots, X_n$ has an extreme value limiting distribution when appropriately normalized.

9.3.2. Example of Inconsistency

Angus (1989) derives the limiting random measure for the maximum and minimum in the case of each of the three extreme value types. He does this by applying the bootstrap principle to the appropriately normalized extremes in a way analogous to what Athreya (1987b) did for the normalized the sample mean in the case of stable laws with $0 < \alpha < 2$.

A key fact that sheds light on the result is that for a fixed integer $r \geq 1$, if we let $X_{(r)}$ denote the rth smallest observation from the sample $X_1, X_2, X_3, \ldots, X_n$, then we have $P[Y_{(1)} = X_{(r)} | X_1, X_2, \ldots, X_n] = P[Y_{(n)} = X_{(n-r+1)} | X_1, X_2, \ldots, X_n] = \left[1 - \frac{(r-1)}{n}\right]^n \left(-1 - \frac{r}{n}\right)^n$, where $Y_{(1)}$ is the minimum of a bootstrap sample $Y_1, Y_2, \ldots, Y_n$ drawn by sampling with replacement from $X_1, X_2, \ldots, X_n$ and $Y_{(n)}$ is the corresponding maximum from that same bootstrap sample. Taking the limit as $n \to \infty$ yields the limiting value $e^{-r}(e - 1)$.

Suppose that a_n^* and b_n^* are the appropriate bootstrap analogs to the normalization constants a_n and b_n for the maximum in Gnedenko's theorem. We would then expect that

$$P\left[\frac{Y_{(n)} - a_n^*}{b_n^*} \le t\right] \xrightarrow{d} H(t, \omega)$$

as $n \to \infty$ where H places probability mass $e^{-r}(e - 1)$ at the random point Z_r for $r = 1, 2, \ldots, n$ where $\xrightarrow{d}$ denotes convergence in distribution and Z_r is determined by the sequence $X_1, X_2, X_3, \ldots$ Then unconditionally

$$(Y_{(n)} - a_n^*)b_n^* \xrightarrow{d} Z$$

where

$$P[Z \le t] = (e-1)\sum_{r=1}^{\infty} e^{-r} P[Z_r \le t].$$

Angus proceeds to determine these conditional random measures and the unconditional distributions in the various cases.

The key point is that $H(t, \omega)$ is again a random probability distribution rather than a fixed distribution. So again the bootstrap principle fails since $(X_{(n)} - a_n)/b_n$ converges in distribution to a fixed extreme value distribution, say G, but $(Y_{(n)} - a_n)/b_n$ converges to a random probability distribution H that necessarily differs from G.

9.3.3. Remedies

The same remedy exists for the extremes as we had for the sample mean in the infinite variance case, namely the m-out-of-n bootstrap with $m = o(n)$. For a proof of this see Fukuchi (1994). Virtues of the m-out-of-n bootstrap have been developed by Bickel, Götze, and van Zwet (1997) and Politis, Romano, and Wolf (1999).

Zelterman (1993) provides a different approach to remedy the problem based on a semiparametric bootstrap. He was the first to find a remedy for the inconsistency of the bootstrap for the minimum and maximum of a sequence of random variables in the IID case.

Hall (1992a) developed a bootstrap theory for functionals that are sufficiently smooth to admit Edgeworth and Cornish–Fisher expansions. These counterexamples obviously fail the smoothness conditions. However, this does not tell the whole story since the smoothness conditions, although sufficient, are far from being necessary conditions for the bootstrap principle to work.

The best that can be said is that there are many practical situations where the bootstrap has been shown to work well empirically, and additionally there is theory which proves that under certain conditions it is guaranteed to work asymptotically. Unfortunately, the counterexamples provided in this chapter show that there are situations where the bootstrap principle fails. This suggests that the practitioner must be very careful when applying the ordinary

bootstrap procedure. In some of these cases there may be modifications that allow a form of bootstrapping to work. This theory needs further development to make the guidelines clear.

9.4. SURVEY SAMPLING

9.4.1. Introduction

In the case of sample surveys, the target population is always finite. If the size of the population is N and the sample size is n and n/N is not very small, then the variance of estimates such as averages is smaller than what it would be based on the usual theory of an infinite population. Recall that for independent identically distributed observations X_i for $i = 1, 2, 3, \ldots, n$, with μ, the population mean and σ^2, the population has the expected value of the sample mean equal to μ and variance of the sample mean equal to σ^2/n.

For a finite population, a random sample of size n taken from a population of size N and also with mean equal to μ and population variance σ^2 also has the expected value of the sample mean equal to μ but the variance of the sample mean equal to $(N - n)\sigma^2/(nN)$. This is smaller than the value in the infinite population by a factor of $(N - n)/N$. This factor is called the finite population correction and the factor $f = n/N$ is called the sampling fraction. Using this notation the sampling fraction is f and the finite population correction is $1 - f$.

In the finite population, we define the population mean and variance as

$$\mu = (X_1 + X_2 + X_3 + \cdots + X_N)/N$$

and

$$\sigma^2 = [(X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_N - \mu)^2]/(N - 1).$$

The reason the variance is smaller is that for finite populations the observations are slightly negatively correlated rather than independent as in the case for the infinite population. To see the correlation relationship, we consider the following algebraic manipulations:

$$\mu = (X_1 + X_2 + X_3 + \cdots + X_N)/N = nX_b/N + (X_{n+1} + X_{n+2} + \cdots + X_N)/N$$

or

$$nX_b/N = [\mu - (X_{n+1} + X_{n+2} + \cdots + X_N)/N,$$
$$SX_b = [\mu - (X_{n+1} + X_{n+2} + X_{n+3} + \cdots + X_N)/N]N/n.$$

See Cochran (1977, pp. 23–24) for a derivation of the variance formula.

9.4.2. Example of Inconsistency

Now if one applies the ordinary bootstrap to the sample mean by sampling with replacement from the n observation in the original sample, the bootstrap will mimic the sampling of independent observations, and consequently the variance in the bootstrap samples of the sample mean will be σ^2/n rather than the correct value $(1 - f)\sigma^2/n$. So in this sense the bootstrap fails.

9.4.3. Remedies

In most cases since f is known, a simple finite sample correction factor can be applied to correct the variance estimate. This is achieved by simply multiplying the bootstrap sample estimate of variance by the factor $1 - f$ as the appropriate adjustment.

Another approach that works here also is to use the m out of n bootstrap. For an appropriate choice of m, this bootstrap will provide consistent estimates. The same problem carries over to other estimators. Of course, if f is very small, say 1–5%, the effect on the variance is small. Davison and Hinkley (1997) point out that for many practical cases, f could be between 10% and 50% and then it matters a great deal.

Stratification is common in survey sampling as estimates by common groups. Also, stratification can be useful in reducing the variability of some estimates.

Stratified bootstrapping is very simple. We just sample with replacement within each strata. Further discussion on bootstrapping from finite populations can be found in Davison and Hinkley (1997, pp. 92–100).

In survey sampling a randomization scheme that samples from groups proportional to a measure of size is sometimes done for convenience and other times for very sound statistical reasons. In either case, variance estimators are required that account for the sampling mechanism. Kaufman (1998) provides a bootstrap estimate for variance in a systematic sampling scheme that sample proportionally to size.

9.5. DATA SEQUENCES THAT ARE *M*-DEPENDENT

9.5.1. Introduction

An infinite sequence of random variables $\{X_j\}$ for $j = 1, 2, 3, 4, \ldots, n, \ldots, \infty$ is called m-dependent if for every j, X_j and X_{j+m-1} are dependent but X_j and X_{j+m} are independent. A common example of this in time series analysis is the mth-order moving average processes given by the following equation:

$$X_j - \mu = \sum_{l=1}^{m} \alpha_l \varepsilon_{j-l} \qquad \text{for } j \in Z,$$

where μ is the mean of the stationary sequence, $\{X_j\}$ and $\{\alpha_l\}$ are the moving average coefficients and the variables $\{\varepsilon_j\}$ for $j = 1, 2, 3, 4, \ldots, n, \ldots \infty$ are called the innovations and are assumed to be IID.

This stationary sequence or time series is m-dependent. The reason for this is that any of the X_j that are separated by less than m time units share at least one innovation in common. However, if they are separated by m or more time units, they have no innovations in common and since the innovations are independent of each other and these Xs are linear combinations of disjoint sets of innovations, they too must be independent of each other.

In the next section, we demonstrate how the bootstrapping the sample mean of an m-dependent sequence using the IID nonparametric bootstrap will be an inconsistent estimate of the mean.

9.5.2. Example of Inconsistency when Independence Is Assumed

In Singh (1981), the first proof [also accomplished in Bickel and Freedman (1981)] of the consistency of the bootstrap for the sample mean in the IID case when certain moment conditions are satisfied was established. In that same seminal paper, Singh provided an example of the inconsistency of the same sample mean when the data are m-dependent rather than IID. In the example, Singh assumed that the m-dependent stationary sequence $\{X_i\}$, for $i = 1, 2, 3, \ldots, n, \ldots$, has $EX_1 = \mu$ and $EX_1^2 < \infty$.

If we let $\sigma_m^2 = \mathrm{Var}(X_1) + 2\sum_{i=1}^{m-1} \mathrm{Cov}(X_1, X_{1+i})$ and $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$.

By the central limit theorem for m-dependent sequences [see, for example, Theorem A.7, Appendix A in Lahiri (2003a)], we have

$$\sqrt{n}(\bar{X}_n - \mu) \to N(0, \sigma_m^2), \qquad \text{where the convergence is in distribution.}$$

Suppose we want to estimate the sampling distribution of the random variable $T_n = \sqrt{n}(\bar{X}_n - \mu)$ using the IID bootstrap. Then the bootstrap version of T_n is given by

$$T_{n,n}^* = \sqrt{n}(\bar{X}_n^* - \bar{X}_n), \qquad \text{where } \bar{X}_n^* = n^{-1}\sum_{i=1}^{n} X_i^*.$$

In this case we assume that the resample size is n, the same as the original size of the sequence. Theorem 2.2, page 21 shows that $T_{n,n}^*$ converges in distribution to $N(0, \sigma^2)$, where $\sigma^2 = \mathrm{Var}(X_1)$. But since $\sigma^2 \neq \sigma_m^2$ the variance is wrong and hence the estimate is not consistent.

9.5.3. Remedies

Since the only problem preventing consistency is that the normalization is wrong, a remedy would be to simply correct the normalization.

9.6. UNSTABLE AUTOREGRESSIVE PROCESSES

9.6.1. Introduction

An unstable autoregressive process of order p is given by the following equation:

$$X_j = \sum_{l=1}^{p} \rho_l X_{j-l} + \varepsilon_j \qquad \text{for } j \in Z,$$

where the variables $\{\varepsilon_j\}$ for $j = 1, 2, 3, 4, \ldots, n, \ldots \infty$ is the innovation series and at least one of the roots of the characteristic polynomial is on the unit circle and the others are inside the unit circle. We shall see in the next section that the autoregressive bootstrap defined in Chapter 5 for stationary autoregressive processes fails to be consistent in the unstable case.

9.6.2. Example of Inconsistency

For the unstable autoregressive process of order p defined in Section 9.6.1 the autoregressive bootstrap (ARB) defined in Chapter 5 for the stationary pth-order autoregressive process (a process with a characteristic polynomial that has all its roots inside the unit circle) is not consistent when applied to an unstable pth-order autoregressive process.

To illustrate the problem, let us consider the simplest case the AR(1) model. In the unstable case this model is given by the following equation:

$$X_j = \rho_1 X_{j-1} + \varepsilon_j,$$

where the variables $\{\varepsilon_j\}$ for $j = 1, 2, 3, 4, \ldots, n, \ldots \infty$ is the innovation series and at least one of the roots of the characteristic polynomial is on the unit circle. Now since this is first order, the characteristic polynomial is linear and hence has only one root, and that root must lie on the unit circle. The characteristic polynomial in this case is $\Psi_1(Z) = Z - \rho_1$. The root of this polynomial is $Z = \rho_1$ for any real value of $\Psi_1(Z) = Z - \beta_1$. The only values that ρ_1 can take on and fall on the unit are -1 and 1. Lahiri points out [Lahiri (2003a, p. 209)] that the least squares estimate of ρ_1 is still a consistent estimate but has a different rate of convergence and a different limiting distribution than for either the stationary or the explosive cases. This helps the reader gain insight as to why the ARB bootstrap would fail to be consistent as an estimate for ρ_1.

For a proof that the ARB bootstrap with resample size $m = n$ is inconsistent, see Datta (1996). Just as in the case of estimating a mean when the variance is infinite and the case of estimating the maximum value of an IID sequence, the limiting distribution for the test statistic that would be used in the stationary case has a random limiting distribution when the process is unstable.

9.6.3. Remedies

Just as in other cases described in this chapter, an m-out-of-n bootstrap provides an easy fix. Datta (1996) and Heimann and Kreiss (1996) independently show that ARB with a resample size m going to infinity at a slower rate than n. The theorems require conditions on the moments of the innovation series. Datta's theorem requires the existence of a $2 + \delta$ moment for the innovations, while the theorem of Heimann and Kress only assumes that the second moment exists. However, Datta proves a stricter rate of convergence (almost sure convergence), whereas Heimann and Kress prove only convergence in probability.

Since the problem with the bootstrap is due to the least-squares estimate of ρ_1 and the estimate of ρ_1 is used to obtain the residuals that are bootstrapped, one idea is to simply modify this estimate. Datta and Sriram (1997) use a shrinkage estimate of ρ_1 in place of the least-squares estimate and since their estimate has a faster rate of convergence, they are able to show that this modified ARB is consistent.

For the m-out-of-n bootstrap there is always an issue of how to choose m in practice for a given n. In a special case, Datta and McCormick (1995a) use a version of the jackknife after bootstrap technique to choose m. The paper by Sakov and Bickel (1999) also addresses the issue of choice of m in the m-out-of-n bootstrap.

9.7. LONG-RANGE DEPENDENCE

9.7.1. Introduction

The stationary ARMA that we have previously considered fall into the category of weakly dependent processes because their autocorrelation function eventually decays exponentially. Stationary stochastic processes that have long-range dependence have the early observations continue to have a large influence on observations much later in time. A mathematical characterization of this is that the sum of the autocorrelations is divergent.

Although we choose this as our definition of long-range dependence, this is not the only way to mathematically characterize this concept. Both Beran (1994) and Hall (1997) consider various possibilities for long-range dependence.

9.7.2. Example of Inconsistency

Lahiri shows with Theorem 10.2 on page 244 of Lahiri (2003a) that the MBB fails to provide a valid approximation to the distribution of the normalized sample mean under long-range dependence. In this case the difficulty is that the rate of convergence of the mean in the weakly dependent case is much

faster than in the long-range dependence case. So the normalization causes the sample mean to converge to a degenerate limit.

9.7.3. A Remedy

The scaling factor must tend to zero at a slower rate in order to get a nondegenerate limit. Lahiri's Theorem 10.3 on page 245 of Lahiri (2003a) shows that the appropriate remedy in order to get convergence to a weak limit is to use the MBB with a modified scaling constant.

9.8. BOOTSTRAP DIAGNOSTICS

In parametric problems, such as linear regression, there are many assumptions that are required to make the methods work well (e.g., for the normal linear models (1) homogeneity of variance, (2) normally distributed residuals, and (3) no correlation or trend among the residuals). These assumptions can be checked using various statistical tests or through numerical or graphical diagnostic tools. Similar diagnostics can be applied in the case of parametric bootstraps.

However, in the case of the nonparametric bootstrap, the number of assumptions that are required is minimal and it is difficult to characterize conditions when the bootstrap can be expected to work or fail. Consequently, it was believed for a long time that no diagnostics could be developed for this bootstrap.

Efron (1992c) introduced the concept of a jackknife-after-bootstrap measure that could be used as a tool for accessing the nonparametric bootstrap. The idea is to illustrate what the effect of leaving out individual observations has on bootstrap calculations. This idea addresses a basic question: Once a bootstrap calculation has been performed, how different might the result have been if a single observation, say y_j, had been left out of the original data?

An intuitive approach might be to do the resampling over again with the original replaced by the data with y_j left out. We would then compare the two bootstrap estimates to determine the effect. However, we really would want to see what the effect is for each observation left out. This would be even more computationally intensive than just bootstrapping because we would be repeating the entire resampling procedure n times!

However, this brute force approach is not necessary because, computationally, we can do something equivalent without repeating the bootstrap sampling. What we do is take advantage of the information in the original bootstrap samples that have the observation y_j left out.

If we did enough bootstrap replications originally (say, 1000–5000), there should be many cases where each observation is left out. In fact for any bootstrap sample there is approximately a probability = $1/e \approx 0.368$ that any

particular observation is not included. You may recall the use of this argument in the heuristic justification for the .632 estimator of error rate in discriminant analysis. So for each observation, there should be a set of approximately 36.8% of the bootstrap samples that are missing this particular observation.

To illustrate, suppose we are estimating the bias of an estimator t. Let B be the bootstrap estimate of the bias based on the entire bootstrap sample. Consider the subset N_j of bootstrap samples that do not contain the observation y_j. Let B_{-j} denote the bootstrap estimate of bias for t based on the same bootstrap calculation using only the subset of N_j of bootstrap samples. The jackknife estimate then scales by the sample size n and provides an estimator of the effect as $n(B_{-j} - B)$.

This can be repeated for all j. Each value is analogous to the pseudo-value in ordinary jackknife estimation. Such an effect is very much akin to empirical influence function estimates. Because this jackknife estimate is applied after bootstrapping, it is called a jackknife-after-bootstrap measure.

In fact, one suggested diagnostic is obtained by plotting these jackknife after-bootstrap measures with empirical influence function values, possibly standardized. More details along with an example can be found in Davison and Hinkley (1997, pp. 114–116).

9.9. HISTORICAL NOTES

Bickel and Freedman (1981) and Knight (1989) consider the example of the bootstrap distribution for the maximum of a sample from the uniform distribution $(0, \theta)$. Since the parameter θ is the upper endpoint of the distribution, the maximum $X_{(n)}$ increases to as $n \to \infty$. This is a case where regularity conditions fail. They show that the bootstrap distribution for the normalized maximum converges to a random probability measure. This is also discussed in Schervish (1995, Example 5.80, page 330).

Consequently, Bickel and Freedman (1981) provide an early counterexample to the correctness of the bootstrap principle. Work by Angus (1993) extends this result to general cases of the limiting distribution for maxima and minima of independent identically distributed sequences.

Angus' work was motivated by the results in Athreya (1987b), who worked out the counterexamples for the sample mean when second moments fail to exist but the limiting stable law is well-defined. Knight (1989) provides an alternative proof of Athreya's result.

A good general account of extreme value theory can be found in Leadbetter, Lindgren, and Rootzen (1983), Galambos (1978, 1987), and Resnick (1987). Reiss (1989) presents the application of the bootstrap distribution for the sample quantiles (Reiss 1989, pp. 220–226). Another good account is given in Reiss and Thomas (1997), who provide both examples and software. They also discuss bootstrap confidence intervals (pp. 82–84). The book edited

by Adler, Feldman, and Taqqu (1998) contains two articles that show ways to bootstrap when distributions gave heavy tails (see Pictet, Dacorogna, and Muller, 1998, pp. 283–310; LePage, Podgórski, Ryznar, and White 1998, pp. 339–358).

A simple illustration of the inconsistency of the minimum is given in Reiss (1989, page 221). Castillo (1988) presents some of the theory along with engineering application. Bootstrapping extremes in dependent cases and bootstrapping parameters in heavy-tailed distribution is covered in Lahiri (2003a, Chapter 11).

Singh (1981) and Bickel and Freedman (1981) were the first to show that the bootstrap principle works for the sample mean when finite second moments exist. This was an important theoretical result that gave Efron's bootstrap a big shot in the arm, since it provided stronger justification than simple heuristics and analogies to the jackknife and other similar methods.

Since then, efforts (Yang, 1988; Angus, 1989) have been made to make the proof of the asymptotic normality of the bootstrap mean in the finite variance case simpler and more self-contained.

The treatise by Mammen (1992b) attempts to show when the bootstrap can be relied on, based on asymptotic theory and simulation results. The bootstrap conference in Ann Arbor, Michigan in 1990 attempted to look at new applications and demonstrate the limitations of the bootstrap. See LePage and Billard (1992) for a collection of papers from that conference.

Combinational results from classical occupancy theory can be found in Feller (1971) and Johnson and Kotz (1977). Chernick and Murthy (1985) apply these results to obtain the repetition frequencies for the bootstrap samples. Hall (1992a, Appendix I) discusses the atoms of the bootstrap distribution. In addition, general combinational results applicable to the bootstrap distribution can be found in Roberts (1984).

Regarding finite populations, we have already mentioned Cochran (1977) as one of the classic texts. Variance estimation by balanced subsampling methods goes back to McCarthy (1969). The first attempt to apply the bootstrap in the finite population setting is Gross (1980). His method is called the "population" bootstrap and was restricted to cases where N/n is an integer. Bickel and Freedman (1984) and Chao and Lo (1994) apply the approach that Davison and Hinkley advocate. Booth, Butler, and Hall (1994) describe the construction of studentized bootstrap confidence limits in the finite population context.

Presnell and Booth (1994) give a critical discussion of earlier literature, and based on the superpopulation model approach to survey sampling they describe the superpopulation bootstrap and modified sample size approach was presented by McCarthy and Snowden (1985), and the mirror-matched method was presented by Sitter (1992a). A rescaling approach was introduced by Rao and Wu (1988).

A comprehensive account of both the jackknife and the bootstrap approaches to survey sampling can be found in Shao and Tu (1995, Chapter

6). Kovar (1985, 1987) and Kovar, Rao, and Wu (1988) performed simulation studies to compare various resampling variance estimators and confidence intervals in the case of stratified one-stage simple random sampling with replacement. Shao and Tu (1995) provide a summary of these studies and their findings. Shao and Sitter (1996) apply the bootstrap to imputation problems in the survey sampling context.

Excellent coverage of the bootstrap for stationary, unstable, and explosive autoregressive processes is given by in Lahiri (2003a). Important work on consistency results can be found in Datta (1996), Datta (1995), Datta and McCormick (1995a), and Datta and Sriram (1997).

As mentioned earlier the jackknife-after-bootstrap diagnostics were introduced by Efron (1992c). Different graphical diagnostics for the reliability of the bootstrap have been developed in an asymptotic framework in Beran (1997). Linear plots for diagnosis of problems with the bootstrap are presented in Davison and Hinkley (1997, page 119). This approach is due to Cook and Weisberg (1994).

Bibliography 1 (Prior to 1999)

Aastveit, A. H. (1990). Use of bootstrapping for estimation of standard deviation and confidence intervals of genetic variance and covariance components. *Biomed. J.* **32**, 515–527.*

Abel, U., and Berger, J. (1986). Comparison of resubstitution, data splitting, the bootstrap and the jackknife as methods for estimating validity indices of new marker tests: A Monte Carlo study. *Biomed. J.* **28**, 899–908.*

Abramovitch, L., and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13**, 116–132.*

Acutis, M., and Lotito, S. (1997). Possible application of resampling methods to statistical analysis of agronomic data. *Riv. Agron.* **31**, 810–816.

Aczel, A. D., Josephy, N. H., and Kunsch, H. R. (1993). Bootstrap estimates of the sample bivariate autocorrelation and partial autocorrelation distributions. *J. Statist. Comput. Simul.* **46**, 235–249.

Adams, D. C., Gurevitch, J., and Rosenberg, M. S. (1997). Resampling tests for meta-analysis of ecological data. *Ecology* **78**, 1277–1283.*

Adkins, L. C. (1990). Small sample performance of jackknife confidence intervals for the James–Stein estimator. *Commun. Statist. Simul. Comput.* **19**, 401–418.

Adkins, L. C., and Hill, R. C. (1990). An improved confidence ellipsoid for the linear regression model. *J. Statist. Comput. Simul.* **36**, 9–18.

Adler, R. J., Feldman, R. E., and Taqqu, M. S. (editors) (1998). *A Practical Guide to Heavy Tails*. Birkhauser, Boston.*

Aebi, M., Embrechts, P., and Mokosch, T. (1994). Stochastic discounting, aggregate claims and the bootstrap. *Adv. Appl. Probab.* **26**, 183–206.

Aegerter, P., Muller, F., Nakache, J. P., and Boue, A. (1994). Evaluation of screening methods for Down's syndrome using bootstrap comparison of ROC curves. *Comput. Methods Prog. Biomed.* **43**, 151–157.*

Aerts, M., and Gijbels, I. (1993). A three-stage procedure based on bootstrap critical points. *Commun. Statist. Seq. Anal.* **12**, 93–113.

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

Aerts. M., Janssen, P., and Veraverbeke, N. (1994). Bootstrapping regression quantiles. *J. Nonparametric Statist.* **4**, 1–20.

Agresti, A. (1990). *Categorical Data Analysis*. Wiley, New York.*

Ahn, H., and Chen, J. J. (1997). Tree-structured logistic model for overdispersed binomial data with application to modeling development effects. *Biometrics* **53**, 435–455.

Aitkin, M., Anderson, D., and Hinde, J. (1981). Statistical modelling of data on teaching styles (with discussion). *J. R. Statist. Soc. A* **144**, 419–461.

Aitkin, M., and Tunnicliffe-Wilson, G. (1980). Mixture models, outliers and the EM algorithm. *Technometrics* **22**, 325–331.*

Akritas, M. G. (1986). Bootstrapping the Kaplan–Meier estimator. *J. Am. Statist. Assoc.* **81**, 1032–1038.*

Alemayehu, D. (1987a). Approximating the distributions of sample eigenvalues. *Proc. Comput. Sci. Statist.* **19**, 535–539.

Alemayehu, D. (1987b). Bootstrap method for multivariate analysis. *Proc. Statist. Comp.* 321–324, American Statistical Association, Alexandria.

Alemayehu, D. (1988). Bootstrapping the latent roots of certain random matrices *Commun. Statist. Simul. Comput.* **17**, 857–869.

Alemayehu, D., and Doksum, K. (1990). Using the bootstrap in correlation analysis with a longitudinal data set. *J. Appl. Statist.* **17**, 357–368.

Alkuzweny, B. M. D., and Anderson, D. A. (1988). A simulation study of bias in estimation of variance by bootstrap linear regression model. *Commun. Statist. Simul. Comput.* **17**, 871–886.

Allen, D. L. (1997). Hypothesis testing using an L_1–distance bootstrap design. *Am. Statist.* **51**, 145–150.

Albanese, M. T., and Knott, M. (1994). Bootstrapping latent variable models for binary response. *Br. J. Math. Statist. Psychol.* **47**, 235–246.

Altman, D. G., and Andersen, P. K. (1989). Bootstrap investigation of the stability of a Cox regression model. *Statist. Med.* **8**, 771–783.*

Amari, S. (1985). *Differential Geometrical Methods in Statistics*. Springer-Verlag, Berlin.*

Ames, G. A., and Muralidhar, K. (1991). Bootstrap confidence intervals for estimating audit value from skewed populations and small samples. *Simulation* **56**, 119–128.*

Anderson, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.*

Anderson, T. W. (1959). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.*

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.*

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed., Wiley, New York.*

Andrade, I., and Proenca, I. (1992). Search for a break in Portuguese GDP 1833–1985 with bootstrap methods. In *Bootstrapping and Related Techniques, Proceedings, Trier, FRG, Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 133–142. Springer-Verlag, Berlin.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton.*

Andrieu, G., Caraux, G., and Gascuel, O. (1997). Confidence intervals of evolutionary distances between sequences and comparison with usual approaches including the bootstrap method. *Mol. Biol. Evol.* **14**, 875–882.*

Angus, J. E. (1989). A note on the central limit theorem for the bootstrap mean. *Commun. Statist. Theory Methods* **18**, 1979–1982.*

Angus, J. E. (1993). Asymptotic theory for bootstrapping the extremes. *Commun. Statist. Theory Methods* **22**, 15–30.*

Angus, J. E. (1994). Bootstrap one-sided confidence intervals for the log-normal mean. *Statistician* **43**, 395–401.

Archer, G., and Chan, K. (1996). Bootstrapping uncertainty in image analysis. In *Proceedings in Computational Statistics, 12th Symposium*, pp. 193–198.

Arcones, M. A., and Gine, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Ann. Inst. Henri Poincare* **25**, 457–481.*

Arcones, M. A., and Gine, E. (1991a). Some bootstrap tests of symmetry for univariate continuous distributions. *Ann. Statist.* **19**, 1496–1511.

Arcones, M. A., and Gine, E. (1991b). Additions and corrections to "The bootstrap of the mean with arbitrary bootstrap sample size." *Ann. Inst. Henri Poincare* **27**, 583–595.

Arcones, M. A., and Gine, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 13–47. Wiley, New York.*

Arcones, M. A., and Gine, E. (1994). *U*-processes indexed Vapnik-Cervonenkis classes of functions with applications to asymptotics and bootstrap of *U*-statistics with estimated parameters. *Stoch. Proc.* **52**, 17–38.

Armstrong, J. S., Brodie, R. J., and McIntyre, S. H. (1987). Forecasting methods for marketing: Review of empirical research. *Int. J. For.* **3**, 355–376.

Arvesen, J. N. (1969). Jackknifing *U*-statistics. *Ann. Math. Statist.* **40**, 2076–2100.

Ashour, S. K., Jones, P. E., and El-Sayed, S. M. (1991). Bootstrap investigation on mixed exponential model using weighted least squares. In *26th Annual Conference on Statistics, Computer Science and Operations Research: Mathematical Statistics*, Vol. 1, pp. 79–103, Cairo University.

Athreya, K. B. (1983). Strong law for the bootstrap. *Statist. Probab. Lett.* **1**, 147–150.*

Athreya, K. B. (1987a). Bootstrap of the mean in the infinite variance case. In *Proceedings of the First World Congress of the Bernoulli Society* (Y. Prohorov, and V. Sazonov, editors). Vol. 2, pp. 95–98. VNU Science Press, The Netherlands.

Athreya, K. B. (1987b). Bootstrap estimation of the mean in the infinite variance case. *Ann. Statist.* **15**, 724–731.*

Athreya, K. B., and Fukuchi, J. I (1997). Confidence intervals for endpoints of a c.d.f. via bootstrap. *J. Statist. Plann. Inf.* **58**, 299–320.

Athreya, K. B., and Fuh, C. D. (1992a). Bootstrapping Markov chains: countable case. *J. Statist. Plann. Inf.* **33**, 311–331.*

Athreya, K. B., and Fuh, C. D. (1992b). Bootstrapping Markov chains. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 49–64. Wiley, New York.*

Athreya, K. B., Ghosh, M., Low, L. Y., and Sen, P. K. (1984). Laws of Large Numbers for bootstrapped *U*-statistics. *J. Statist. Plann. Inf.* **9**, 185–194.

Athreya, K. B., Lahari, S. N., and Wei, W. (1998). Inference for heavy-tailed distributions. *J. Statist. Plann. Inf.* **66**, 61–75.

Atwood, C. L. (1984). Approximate tolerance intervals based on maximum likelihood estimates. *J. Am. Statist. Assoc.* **79**, 459–465.

Atzinger, E., Brooks, W., Chernick, M. R., Elsner, B., and Foster, W. (1972). A Compendium on Risk Analysis Techniques. U. S. Army Materiel Systems Analysis Activity, Special Publication No.4.*

Azzalini, A., Bowman, A. W., Hardle, W. (1989). Nonparametric regression for model checking. *Biometrika* **76**, 1–11.

Babu, G. J. (1984). Bootstrapping statistics with linear combination of chi-squares as weak limit. *Sankhya A* **46**, 85–93.

Babu, G. J. (1986). A note on bootstrapping the variance of the sample quantiles. *Ann. Inst. Statist. Math.* **38**, 439–443.

Babu, G. J. (1989). Applications of Edgeworth expansions to bootstrap: A review. In *Statistical Data Analysis and Inference* (Y. Dodge, editor), pp. 223–237. Elsevier Science Publishers, Amsterdam.

Babu, G. J. (1992). Subsample and half-sample methods. *Ann. Inst. Statist. Math.* **44**, 703–720.*

Babu, G. J. (1995). Bootstrap for nonstandard cases. *J. Statist. Plann. Inf.* **43**, 197–203.

Babu, G. J. (1998). Breakdown theory for estimators based on bootstrap and other resampling schemes. Preprint, Department of Statistics, Pennsylvania State University.*

Babu, G. J., and Bai, Z. D. (1996) Mixtures of global and local Edgeworth expansions and their applications. *J. Multivar. Anal.* **59**, 282–307.

Babu, G. J., and Bose, A. (1989). Bootstrap confidence intervals. *Statist. Probab. Lett.* **7**, 151–160.*

Babu, G. J., and Feigelson, E. (1996). *Astrostatistics*. Chapman & Hall, New York.*

Babu, G. J., Pathak, P. K., and Rao, C. R. (1992). A note of Edgeworth expansion for the ratio of sample means. *Sankhya A* **54**, 309–322.

Babu, G. J., Pathak, P. K., and Rao, C. R. (1998). Second order correction of the sequential bootstrap. Prepint, Department of Statistics, Pennsylvania State University.*

Babu, G. J., and Rao, C. R. (1993). Bootstrap methodology. In *Handbook of Statistics* (C. R. Rao, editor), Vol. 9, pp. 637–659. North-Holland, Amsterdam.*

Babu, G. J., Rao, C. R., and Rao, M. B. (1992). Nonparametric estimation of specific occurrence/exposure rate in risk and survival analysis. *J. Am. Statist. Assoc.* **87**, 84–89.

Babu, G. J., and Singh, K. (1983). Nonparametric inference on means using the bootstrap. *Ann. Statist.* **11**, 999–1003.*

Babu, G. J., and Singh, K. (1984a). On one term Edgeworth correction by Efron's bootstrap. *Sankhya A* **46**, 195–206.*

Babu, G. J., and Singh, K. (1984b). Asymptotic representations related to jackknifing and bootstrapping L-statistics. *Sankhya A* **46**, 219–232.*

Babu, G. J., and Singh, K. (1985). Edgeworth expansions for sampling without replacement from finite populations. *J. Multivar. Anal.* **17**, 261–278.*

Babu, G. J., and Singh, K. (1989). On Edgeworth expansions in the mixture cases. *Ann. Statist.* **17**, 443–447.*

Bahadur, R., and Savage, L. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27**, 1115–1122.*

Bai, C. (1988). Asymptotic properties of some sample reuse methods for prediction and classification. Ph.D. dissertation: Department of Mathematics, University of California, San Diego.

Bai, C., Bickel, P. J., and Olshen, R. A. (1990). Hyperaccuracy of bootstrap-based prediction. In *Probability in Banach Spaces: Proceedings of the Seventh International Conference* (E. Eberlein, J. Kuelbs, and M. B. Marcus, editors), pp. 31–42. Birkhauser, Boston.

Bai, Z., and Rao, C. R. (1991). Edgeworth expansion of a function of sample means. *Ann. Statist.* **19**, 1285–1315.*

Bai, Z., and Rao, C. R. (1992). A note on Edgeworth expansion for ratio of sample means. *Sankhya A* **54**, 309–322.*

Bai, Z., and Zhao, L. (1986). Edgeworth expansions of distribution function of independent random variables. *Sci. Sin. A* **29**, 1–22.

Bailer, A. J., and Oris, J. T. (1994). Assessing toxicity of pollutants in aquatic systems. In *Case Studies in Biometry* (N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest, and J. Greenhouse, editors), pp. 25–40. Wiley, New York.*

Bailey, R. A., Harding, S. A., and Smith, G. L. (1989). Cross validation. *Encyclopedia of Statistical Sciences*, Supplemental Volume, pp. 39–44. Wiley, New York.

Bailey, W. A. (1992). Bootstrapping for order statistics sans random numbers (operational bootstrapping). In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 309–318. Wiley, New York.*

Bajgier, S. M. (1992). The use of bootstrapping to construct limits on control charts. *Proc. Dec. Sci. Inst.* 1611–1613.*

Baker, S. G., and Chu, K. C. (1990). Evaluating screening for the early detection and treatment of cancer without using a randomized control group. *J. Am. Statist. Assoc.* **85**, 321–327.*

Banks, D. L. (1988). Histospline smoothing the Bayesian bootstrap. *Biometrika* **75**, 673–684.*

Banks, D. L. (1989). Bootstrapping II. *Encyclopedia of Statistical Sciences*, Supplemental Volume, pp. 17–22. Wiley, New York.*

Banks, D. L. (1993). Book Review of *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors). *J. Am. Statist. Assoc.* **88**, 708–710.

Barabas, B., Csorgo, M. J., Horvath, L., and Yandell, B. S. (1986). Bootstrapped confidence bands for percentile lifetime. *Ann. Inst. Statist. Math.* **38**, 429–438.

Barbe, P., and Bertail, P. (1995). *The Weighted Bootstrap*. Lecture Notes in Statistics. Springer-Verlag, New York.*

Barlow, W. E., and Sun, W. H. (1989). Bootstrapped confidence intervals for linear relative risk models. *Statist. Med.* **8**, 927–935.*

Barnard, G. (1963). Comment on "The spectral analysis of point processes" by M. S. Bartlett. *J. R. Statist, Soc. B* **25**, 294.*

Bar-Ness, Y., and Punt, J. B. (1996). Adaptive "bootstrap" CDMA multi-user detector. *Wireless Pers. Commun.* **3**, 55–71.*

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on standardized signed log likelihood ratio. *Biometrika* **73**, 307–322.

Barndorff-Nielsen, O. E., and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman & Hall, London.*

Barndorff-Nielsen, O. E., and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman & Hall, London.*

Barndorff-Nielsen, O. E., and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75**, 374–378.

Barndorff-Nielsen, O. E., James, I. R., and Leigh, G. M. (1989). A note on a semiparametric estimation of mortality. *Biometrika* **76**, 803–805.

Barnett, V., and Lewis, T. (1995). *Outliers in Statistical Data*, 3rd ed. Wiley, Chichester.*

Barraquand, J. (1995). Monte Carlo integration, quadratic resampling and asset pricing. *Math. Comput. Simul.* **38**, 173–182.

Bartlett, M. S. (1975). *The Statistical Analysis of Spatial Pattern*. Chapman & Hall, London.*

Basawa, I. V., Green, T. A., McCormick, W. P., and Taylor, R. L. (1990). Asymptotic bootstrap validity for finite Markov chains. *Commun. Statist. Theory Methods* **19**, 1493–1510.*

Basawa, I. V., Mallik, A. K., McCormick, W. P., and Taylor, R. L. (1989). Bootstrapping explosive first order autoregressive processes. *Ann. Statist.* **17**, 1479–1486.*

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., and Taylor, R. L. (1991a). Bootstrapping unstable first order autoregressive processes. *Ann. Statist.* **19**, 1098–1101.*

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., and Taylor, R. L. (1991b). Bootstrap test of significance and sequential bootstrap estimators for unstable first order autoregressive processes. *Commun. Statist. Theory Methods* **20**, 1015–1026.*

Basford, K. E., and McLachlan, G. J. (1985a). Estimation of allocation rates in a cluster analysis context. *J. Am. Statist. Assoc.* **80**, 286–293.*

Basford, K. E., and McLachlan, G. J. (1985b). Cluster analysis in randomized complete block design. *Commun. Statist. Theory Methods* **14**, 451–463.*

Basford, K. E., and McLachlan, G. J. (1985c). The mixture method of clustering applied to three-way data. *Classification* **2**, 109–125.*

Basford, K. E., and McLachlan, G. J. (1985d). Likelihood estimation with normal mixture models. *Appl. Statist.* **34**, 282–289.*

Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. Wiley, New York.*

Bau, G. J. (1984). Bootstrapping statistics with linear combinations. *Sankhya A* **46**, 195–206.*

Bauer, P. (1994). Book Review of *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (by P. Westfall and S. S. Young). *Statist. Med.* **13**, 1084–1086.

Beadle, E. R., and Djuric, P. M. (1997). A fast weighted Bayesian bootstrap filter for nonlinear modelstate estimation. *IEEE Trans. Aerosp. Electron. Syst.* **33**, 338–343.

Bean, N. G. (1995). Dynamic effective bandwidths using network observation and the bootstrap. *Aust. Telecommun. Res.* **29**, 43–52.

Bedrick, E. J., and Hill, J. R. (1992). A generalized bootstrap. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 319–326. Wiley, New York.*

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data Sources of Collinearity*. Wiley, New York.*

Belyaev, Y. K. (1996). Resampling from realizations of random processes and fields, with application to statistical inference. In *4th World Congress of the Bernoulli Society, Vienna*.

Benichou, J., Byrne, C., and Gail, M. (1997). An approach to estimating exposure-specific rate of breast cancer from a two-stage case-control study within a cohort. *Statist. Med.* **16**, 133–151.

Bensmail, H., and Celeux, G. (1996). Regularized Gaussian discriminant analysis through eigenvalue decomposition. *J. Am. Statist. Assoc.* **91**, 1743–1749.

Bentler, P. M., and Yung, Y.-F. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. *Br. J. Math. Statist. Psychol.* **47**, 63–84.

Beran, J. (1994). *Statistics for Long Memory Processes*. Chapman & Hall, London.*

Beran, R. J. (1982). Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.* **10**, 212–225.*

Beran, R. J. (1984a). Jackknife approximations to bootstrap estimates. *Ann. Statist.* **12**, 101–118.*

Beran, R. J. (1984b). Bootstrap methods in statistics. *Jahreshber. D. Dt. Math. Verein.* **86**, 14–30.*

Beran, R. J. (1985). Stochastic procedures: bootstrap and random search methods in statistics. *Proceedings of the 45th Session of the ISI*, **4**, 25.1, Amsterdam.

Beran, R. J. (1986). Simulated power functions. *Ann. Statist.* **14**, 151–173.*

Beran, R. J. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457–468.*

Beran, R. J. (1988a). Weak convergence: statistical applications. *Encyclopedia of Statistical Sciences*, Vol. 9, pp. 537–539. Wiley, New York.

Beran, R. J. (1988b). Balanced simultaneous confidence sets. *J. Am. Statist. Assoc.* **83**, 679–686.

Beran, R. J. (1988c). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *J. Am. Statist. Assoc.* **83**, 687–697.*

Beran, R. J. (1990a). Refining bootstrap simultaneous confidence sets. *J. Am. Statist. Assoc.* **85**, 417–426.*

Beran, R. J. (1990b). Calibrating pediction regions. *J. Am. Statist. Assoc.* **85**, 715–723.*

Beran, R. J. (1992). Designing bootstrap predictions. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG* (K.-H. Jockel, G., and W. Sendler, editors). *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 23–30. Springer-Verlag Berlin.*

Beran, R. J. (1993). Book Review of *The Bootstrap and Edgeworth Expansion* (by P. Hall). *J. Am. Statist. Assoc.* **88**, 375–376.

Beran, R. J. (1994). Seven stages of bootstrap. In *25th Conference on Statistical Computing: Computational Statistics* (P. Dirschdl, and R. Ostermann, editors), pp. 143–158. Physica-Verlag, Heidelberg.*

Beran, R. J. (1996). Bootstrap variable-selection and confidence sets. In *Robust Statistics, Data Analysis and Computer Intensive Methods* (P. J. Huber, and H. Rieder, editors). *Lecture Notes in Statistics*, Vol. 109, pp. 1–16. Springer-Verlag, New York.

Beran, R. J. (1997). Diagnosing bootstrap success. *Ann. Inst. Statist. Math.* **49**, 1–24.*

Beran, R. J., and Ducharme, G. R. (1991). *Asymptotic Theory for Bootstrap Methods in Statistics*. Les Publications Centre de Recherches Mathematiques, Universite de Montreal, Montreal.*

Beran, R. J., and Fisher, N. L. (1998). Nonparametric comparison of mean directions or mean axes. *Ann. Statist.* **26**, 472–493.

Beran, R. J., LeCam, L., and Millar, P. W. (1985). Asymptotic theory of confidence sets. In *Proceedings Berkeley Conference in Honor of J. Neyman and J. Kiefer* (L. M. LeCam, and R. A. Olshen, editors), Vol. 2, pp. 865–887, Wadsworth, Monterey.

Beran, R. J., and Millar, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* **14**, 431–443.

Beran, R. J., and Millar, P. W. (1987). Stochastic estimation and testing. *Ann. Statist.* **15**, 1131–1154.

Beran, R. J., and Millar, P. W. (1989). A stochastic minimum distance test for multivariate parametric models. *Ann. Statist.* **17**, 125–140.

Beran, R. J., and Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix, *Ann. Statist.* **13**, 95–115.*

Beran, R. J., and Srivastava, M. S. (1987). Correction to Bootstrap tests and confidence regions for functions of a covariance matrix, *Ann. Statist.* **15**, 470–471.*

Bernard, V. L. (1987). Cross-sectional dependence and problems in inference in market-based accounting. *J. Acct. Res.* **25**, 1–48.

Bernard, J. T., and Veall, M. R. (1987). The probability distribution of future demand: the case of hydro Quebec. *J. Bus. Econ. Statist*, **5**, 417–424.

Bertail, P. (1992). Le method du bootstrap quelques applications et results theoretiques. Ph.D. dissertation. University of Paris IX.

Bertail, P. (1997). Second-order properties of an extrapolated bootstrap without replacement under weak assumptions. *Bernoulli* **3**, 149–179.

Besag, J. E., and Diggle, P. J. (1977). Simple Monte Carlo tests for spatial patterns. *Appl. Statist.* **26**, 327–333.*

Besag, J. E., and Clifford, P. (1989). Generalized Monte Carlo significance tests. *Biometrika* **76**, 633–642.

Besag, J. E., and Clifford, P. (1991). Sequential Monte Carlo *p*-values. *Biometrika* **78**, 301–304.

Besse, P., and de Falguerolles, A. (1993). Application of resampling methods to the choice of dimension in principal component analysis. In *Computer Intensive Methods in Statistics*, pp. 167–176. Physica-Verlag, Heidelberg, Germany, W. Härdlegl, L. Simar editors.

Bhatia, V. K., Jayasankar, J., and Wahi, S. D. (1994). Use of bootstrap technique for variance estimation of heritability estimators. *Ann. Agric. Res.* **15**, 476–480.

Bhattacharya, R. N. (1987). Some aspects of Edgeworth expansions in statistics and probability. In *New Perspectives in Theoretical and Applied Statistics* (M. Puri, J. P. Vilaplana, and W. Wertz, editors), pp. 157–170. Wiley, New York.

Bhattacharya, R. N., and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 435–451.*

Bhattacharya, R. N., and Qumsiyeh, M. (1989). Second-order and L^p comparisons between the bootstrap and empirical Edgeworth expansion methodologies. *Ann. Statist.* **17**, 160–169.*

Bianchi, C., Calzolari, G., and Brillet, J.-L. (1987). Measuring forecast uncertainty: a review with evaluation based on a macro model of the French economy. *Int. J. For.* **3**, 211–227.

Bickel, P. J. (1992). Theoretical comparison of different bootstrap-*t* confidence bounds. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 65–76. Wiley, New York.*

Bickel, P. J., and Freedman, D. A. (1980). On Edgeworth expansions for the bootstrap. Technical Report. Department of Statistics, University of California, Berkeley.

Bickel, P. J., and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196–1217.*

Bickel, P. J., and Freedman, D. A. (1983). Bootstrapping regression models with many parameters. In *Festschrift for Erich Lehmann* (P. J. Bickel, K. Doksum, and J. L. Hodges, editors), pp. 28–48. Wadsworth, Belmont, CA.*

Bickel, P. J., and Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12**, 470–482.*

Bickel, P. J., and Ghosh, J. K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction. A Bayesian argument. *Ann. Statist.* **18**, 1070–1090.

Bickel, P. J., Götze, F., and van Zwet, W. R. (1997). Resampling fewer than *n* observations, gains, losses, and remedies for losses. *Statist. Sin.* **7**, 1–32.*

Bickel, P. J., Klassen, C. A. J., Ritov,Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore.

Bickel, P. J., and Krieger, A. M. (1989). Confidence bands for a distribution function using bootstrap. *J. Am. Statist. Assoc.* **84**, 95–100.*

Bickel, P. J., and Ren, J.-J. (1996). The *m* out of *n* bootstrap and goodness fit of tests with doubly censored data. In *Robust Statistics, Data Analysis, and Computer-Intensive Methods* (P. J. Huber, and H. Rieder, editors), Lecture Notes in Statistics, Vol. 109, pp. 35–48. Springer-Verlag, New York.*

Bickel, P. J., and Yahav, J. A. (1988). Richardson extrapolation and the bootstrap. *J. Am. Statist. Assoc.* **83**, 387–393.*

Biddle, G., Bruton, C., and Siegel, A. (1990). Computer-intensive methods in auditing bootstrap difference and ratio estimation. Auditing: *J. Practice Theory* **9**, 92–114.*

Biometrics Editors (1997). Book Review of *Modern Digital Simulation Methodology, 11: Univariate and Bivariate Distribution Fitting, Bootstrap Methods & Applications* (E. J. Dudewicz, editor). *Biometrics* **53**, 1564.

Bliese, P., Halverson, R., and Rothberg, J. (1994). Within-group agreement scores using resampling procedures to estimate expected variance. In *Academy of Management Best Papers Proceedings*, pp. 303–307.

Bloch, D. A. (1997). Comparing two diagnostic tests against the same "gold standard" in the same sample. *Biometrics* **53**, 73–85.

Bloch, D. A., and Silverman, B. W. (1997). Monotone discriminant functions and their application in rheumatology. *J. Am. Statist. Assoc.* **92**, 144–153.

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. Wiley, New York.*

Bogdanov, Y. I., and Bogdanova, N. A. (1997). Bootstrap, data structures, and process control in microdynamics. *Russ. Microelectron.* **26**, 155–158.

Bollen, K. A., and Stine, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In *Testing Structural Equation Models* (K. A. Bollen, and J. S. Long, editors), pp. 111–115. Sage Publications, Beverly Hills, CA.*

Bolviken, E., and Skovlund, E. (1996). Confidence intervals for Monte Carlo tests. *J. Am. Statist. Assoc.* **91**, 1071–1078.

Bonate, P. L. (1993). Approximate confidence intervals in calibration using the bootstrap. *Anal. Chem.* **65**, 1367–1372.

Bondesson, L., and Holm, S. (1985). Bootstrap-estimation of the mean square error of the ratio estimatr for sampling without replacement. In *Contributions to Probability and Statistics in Honor of Gunnar Blom* (J. Lanke, and G. Lindgren, editors), pp. 85–96. Studentlitteratur, Lund.

Bone, P. F., Sharma, S., and Shimp, T. A. (1987). A bootstrap procedure for evaluating goodness-of-fit indices of structural equation and confirmatory factor models. *J. Marketing Res.* **26**, 105–111.

Boomsma, A. (1986). On the use of bootstrap and jackknife in covariance structure analysis. In *Compstat 1986*. 205–210 (F. De Antoni, N. Lauro, and A. Rizzi, editors). Physica-Verlag, Heidelberg.

Boos, D. D. (1980). A new method for constructing approximate confidence intervals from *M*-estimates. *J. Am. Statist. Assoc.* **75**, 142–145.

Boos, D. D., and Brownie, C. (1989). Bootstrap methods for testing homogeneous variances. *Technometrics* **31**, 69–82.

Boos, D. D., Janssen, P., and Veraverbeke, N. (1989). Resampling from centered data in the two-sample problem. *J. Statist. Plann. Inf.* **21**, 327–345.

Boos, D. D., and Monahan, J. F. (1986). Bootstrap methods using prior information. *Biometrika* **73**, 77–83.

Booth, J. G. (1994). Book Review of *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (by P. Westfall and S. S. Young). *J. Am. Statist. Assoc.* **89**, 354–355.

Booth, J. G. (1996). Bootstrap methods for generalized linear mixed models with applications to small area estimation. In *Statistical Modelling* (G. U. H. Seeber, B. J. Francis, R. Hatzinger, and G. Steckelberger, editors). *Lecture Notes in Statistics*, Vol. 104, pp. 43–51. Springer-Verlag, New York.

Booth, J. G., and Butler, R. W. (1990). Randomization distributions and saddlepoint approximations in generalized linear models. *Biometrika* **77**, 787–796.

Booth, J. G., Butler, R. W., and Hall, P. (1994). Bootstrap methods for finite populations. *J. Am. Statist. Assoc.* **89**, 1282–1289.*

Booth, J. G., and Do, K.-A. (1994). Automatic importance sampling for the bootstrap. In *7th Conference on the Scientific Use of Statistical Software. SoftStat 93* (F. Faulbaum, editor), pp. 519–526.

Booth, J. G., and Hall, P. (1993a). An improvement on the jackknife distribution function estimator. *Ann. Statist.* **21**, 1476–1485.

Booth, J. G., and Hall, P. (1993b). Bootstrap confidence regions for functional relationships in error-in-variables models. *Ann. Statist.* **21**, 1780–1791.

Booth, J. G., and Hall, P. (1994). Monte-Carlo approximation and the iterated bootstrap. *Biometrika* **81**, 331–340.

Booth, J. G., Hall, P., and Wood, A. T. A. (1992). Bootstrap estimation of conditional distributions. *Ann. Statist.* **20**, 1594–1610.

Booth, J. G., Hall, P., and Wood, A. T. A. (1993). Balanced importance sampling for the bootstrap. *Ann. Statist.* **21**, 286–298.*

Booth, J. G., and Sarkar, S. (1998). Monte Carlo approximation of bootstrap variances. *Am. Statist.* **52**, 354–357.*

Borowiak, D. (1983). A multiple model discrimination procedure. *Commun. Statist. Theory Methods* **12**, 2911–2921.

Borchers, D. L. (1996). Line transect abundance estimation with uncertain detection on the trackline. Ph.D. thesis. University of Capetown, South Africa.

Borchers, D. L., Buckland, S. T., Goedhart, P. W., Clarke, E. D., and Hedley, S. L. (1998). Horvitz-Thompson estimators for double-platform line transect surveys. *Biometrics* **54**, 1221–1237.

Borrello, G. M., and Thompson, B. (1989). A replication bootstrap analysis of the structure underlying perceptions of stereotypic love. *J. Gen. Psychol.* **116**, 317–327.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.* **16**, 1709–1726.*

Bose, A. (1990). Bootstrap in moving average models. *Ann. Inst. Statist. Math.* **42**, 753–768.*

Bose, A., and Babu, G. J. (1991). Accuracy of the bootstrap approximation. *Probab. Theory Relat. Fields* **90**, 301–316.

Boukai, B. (1993). A nonparametric bootstrapped estimate of the changepoint. *J. Nonparametric Statist.* **3**, 123–134.

Box, G. E. P., and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco.*

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, 2nd ed. Holden Day, San Francisco.*

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd ed. Prentice Hall, Engelwood Cliffs, NJ.*

Bradley, D. W. (1985). The effects of visibility bias on time-budget estimates of niche breadth and overlap. *Auk* **102**, 493–499.

Bratley, P., Fox, B. L., and Schrage, L. E. (1987). *A Guide to Simulation*, 2nd ed. Springer-Verlag, New York.*

Braun, W. J., and Kulperger, P. J. (1997). Properties of a Fourier bootstrap method for time series. *Commun. Statist. Theory Methods* **26**, 1326–1336.*

Breidt, F. J., Davis, R. A., and Dunsmuir, W. T. M. (1995). Improved bootstrap prediction intervals for autoregressions. *J. Time Series Anal.* **16**, 177–200.

Breiman, L. (1988). Submodel selection and evaluation in autoregression: The conditional case and little bootstrap. Technical Report 169, Department of Statistics, University of California, Berkeley.

Breiman, L. (1992). The little bootstrap and other methods for dimensionality selection in regression: *x*-fixed prediction error. *J. Am. Statist. Assoc.* **87**, 738–754.*

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics* **38**, 170–177.

Breiman, L. (1996). Bagging predictors. *Mach. Learn.* **24**, 123–140.

Breiman, L., and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Am. Statist. Assoc.* **80**, 580–619.*

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA.*

Breiman, L., and Spector, P. (1992). Submodel selection and evaluation in regression: *X*-random case. *Int. Statist. Rev.* **60**, 291–319.

Breitung, J. (1992). Nonparametric bootstrap tests: some applications. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG* (K.-H. Jockel, G. Rothe, and W. Sendler, editors). *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 87–97. Springer-Verlag, Berlin.

Bremaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York.*

Brennan, T. F., and Milenkovic, P. H. (1995). Fast minimum variance resampling. *Conference Proceedings, Trier, FRG* (K.-H. Jockel, G. Rothe, and W. Sendler, editors). *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 87–97. Springer-Verlag, Berlin.

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionelles. *Ann. Inst. Henri Poincare Sec. B* **19**, 281–296.

Brey, T. (1990). Confidence limits for secondary prediction estimates: Application of the bootstrap to the increment summation method. *Mar. Biol.* **106**, 503–508.*

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, Expanded Edition, Holt Reinhart, and Winston, New York.*

Brockwell, P. J., and Davis, R. A. (1991). *Time Series Methods*, 2nd ed. Springer-Verlag, New York.*

Brostrom, G. (1997). A martingale approach to the changepoint problem. *J. Am. Statist. Assoc.* **92**, 1177–1183.

Brown, J. K. M. (1994). Bootstrap hypothesis tests for evolutionary trees and other dendrograms. *Proc. Natl. Acad. Sci.* **91**, 12293–12297.

Brownstone, D. (1992). Bootstrapping admissible linear model selection procedures. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 327–344. Wiley, New York.*

Brumback, B. A., and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion), *J. Am. Statist. Assoc.* **93**, 961–994.

Bryand, J., and Day, R. (1991). Empirical Bayes analysis for systems of mixed models with linked autocorrelated random effects. *J. Am. Statist. Assoc.* **86**, 1007–1012.

Buckland, S. T. (1980). A modified analysis of the Jolly–Seber capture recapture model. *Biometrics* **36**, 419–435.*

Buckland, S. T. (1983). Monte Carlo methods for confidence interval estimation using the bootstrap technique. *BIAS* **10**, 194–212.*

Buckland, S. T. (1984). Monte Carlo confidence intervals. *Biometrics* **40**, 811–817.*

Buckland, S. T. (1985). Calculation of Monte Carlo confidence intervals. *Appl. Statist.* **34**, 296–301.*

Buckland, S. T. (1993). Book Review of *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors). *J. Appl. Statist.* **20**, 332–333.

Buckland, S. T. (1994a). Book Review of *Computer Intensive Statistical Methods: Validation, Model Validation and Bootstrap* (by J. S. U. Hjorth). *Biometrics* **50**, 586–587.

Buckland, S. T. (1994b). Book Review of *An Introduction to the Bootstrap* (by B. Efron and R. J. Tibshirani). *Biometrics* **50**, 890–891.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53**, 603–618.

Buckland, S. T., and Garthwaite, P. H. (1990). Algorithm AS 259: Estimating confidence intervals by the Robbins–Monro search process. *Appl. Statist.* **39**, 413–424.

Buckland, S. T., and Garthwaite, P. H. (1991). Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics* **47**, 255–268.

Bühlmann, P. (1994). Blockwise bootstrap empirical process for stationary sequences. *Ann. Statist.* **22**, 995–1012.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3**, 123–148.*

Bühlmann, P., and Künsch, H. R. (1995). The blockwise bootstrap for general parameters of a stationary time series. *Scand. J. Statist.* **22**, 35–54.*

Bull, S. B., and Greenwood, C. M. T. (1997). Jackknife bias reduction for polychotomous logistic regression. *Statist. Med.* **16**, 545–560.

Bunke, O., and Droge, B. (1984). Bootstrap and cross-validation estimates of the prediction error for linear regression models. *Ann. Statist.* **12**, 1400–1424.

Bunke, O., and Riemer, S. (1983). A note on the bootstrap and other empirical procedures for testing linear hypotheses without normality. *Statistics* **14**, 517–526.

Bunt, M., Koch, I., and Pope, A. (1995). How many and where are they? A bootstrap-based segmentation strategy for estimating discontinuous functions. In *Conference Proceedings DICTA-95*, pp. 110–115.

Burke, M. D., and Gombay, E. (1988). On goodness-of-fit and the bootstrap. *Statist. Probab. Lett.* **6**, 287–293.

Burke, M. D., and Horvath, L. (1986). Estimation of influence functions. *Statist. Probab. Lett.* **4**, 81–85.

Burke, M. D., and Yuen, K. C. (1995). Goodness-of-fit tests for the Cox model via bootstrap method. *Statist. Probab. Lett.* **4**, 237–256.

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *J. Am. Statist. Assoc.* **89**, 1290–1302.*

Burr, D., and Doss, H. (1993). Confidence bands for the median survival time as a function of the covariates in the Cox model. *J. Am. Statist. Assoc.* **88**, 1330–1340.

Butler, R., and Rothman, E. D. (1980). Prediction intervals based on reuse of the sample. *J. Am. Statist. Assoc.* **75**, 881–889.

Buzas, J. S. (1997). Fast estimators of the jackknife. *Am. Statist.* **51**, 235–240.

Caers, J., Beirlant, J., and Vynckier, P. (1997). Bootstrap confidence intervals for tail indices. *Comput. Statist. Data Anal.* **26**, 259–277.

Cammarano, P., Palm P., Ceccarelli E., and Creti R. (1992). Bootstrap probability that archaea are monophyletic. In *Probability and Bayesian Statistics in Medicine and Biology* (L. Barrai, G. Coletti, and M. Di Bacco, editors). Applied Mathematics Monographs, No. 4, pp. 12–335, Giardini, Pisa.

Campbell, G. (1994). Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statist. Med.* **13**, 499–508.

Cano, R. (1992). On the Bayesian bootstrap. In *Bootstrapping and Related Techniques Proceedings, Trier, FRG* (K.-H. Jockel, G. Rothe, and W. Sendler, editors). *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 159–161. Springer-Verlag, Berlin.

Canty, A. J., and Davison, A. C. (1997). Implementation of saddlepoint approximations in bootstrap distributions. In *The 28th Symposium on the Interface Between Computer Science and Statistics* (L. Billard, and N. I. Fisher, editors), Vol. 28, pp. 248–253. Springer-Verlag, New York.

Cao-Abad, R. (1989). On wild bootstrap confidence intervals. Unpublished manuscript.

Cao-Abad, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Ann. Statist.* **19**, 2226–2231.*

Cao-Abad, R., and Gonzalez-Manteiga, W. (1993). Bootstrap methods in regression smoothing. *J. Nonparametric Statist.* **2**, 379–388.

Carlin, B. P., and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.*

Carlin, B. P., and Gelfand, A. E. (1990). Approaches for empirical Bayes confidence intervals. *J. Am. Statist. Assoc.* **85**, 105–114.

Carlin, B. P., and Gelfand, A. E. (1991). A sample reuse method for accurate parametric empirical Bayes confidence intervals. *J. R. Statist. Soc. B* **53**, 189–200.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14**, 1171–1194.*

Carlstein, E. (1988a). Typical values. *Encyclopedia of Statistical Sciences*, Vol. 9, pp. 375–377, Wiley, New York.

Carlstein, E. (1988b). Bootstrapping ARMA models: Some simulations. *IEEE Trans. Syst.* **16**, 294, 299.

Carlstein, E. (1992). Resampling techniques for stationary time series: Some recent developments. In *IMA Volumes in Mathematics and Their Applications*. *New Directions in Time Series Analysis*. Springer-Verlag, New York.

Carlstein, E., Do, K. A., Hall, P., Hesterberg, T, and Kunsch, H. R. (1998). Match-block bootstrap for dependent data. *Bernoulli* **4**, 305–328.

Carpenter, J. R. (1996). Simulated confidence regions for parameters in epidemiological models. Ph.D. thesis, Department of Statistics, Oxford University.

Carroll, R. J. (1979). On estimating variances of robust estimates when the errors are asymmetric. *J. Am. Statist. Assoc.* **74**, 674–679.

Carroll, R. J., Kuchenhoff, H., Lombard, F., and Stefanski, L. A. (1996). Asymptotics for the SIMEX estimator in nonlinear measurement error models. *J. Am. Statist. Assoc.* **91**, 242–250.

Carroll, R. J., and Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman & Hall, New York.*

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall, New York.*

Carson, R. T. (1985). SAS macros for bootstrapping and cross-validation regression equations. *SAS SUGI* **10**, 1064–1069.

Carter, E. M., and Hubert, J. J. (1985). Analysis of parallel-line assays with multivariate responses. *Biometrics* **41**, 703–710.

Castillo, E. (1988). *Extreme Value Theory in Engineering*. Academic Press, New York.*

Castillo, E., and Hadi, A. S. (1995). Modeling lifetime data with applications to fatigue models. *J. Am. Statist. Assoc.* **90**, 1041–1054.

Chambers. J., and Hastie, T. J. (editors) (1991). *Statistical Models in S*. Wadsworth, Belmont, CA.*

Chan, E. (1996). An application of a bootstrap to the Lisrel modeling of two-wave psychological data. In *1st Conference on Applied Statistical Science* (M. Ahsanullah, and D. S. Bhoj, editors), Vol. 1, pp. 165–174. Nova Science, Commack, NY.

Chan,Y. M., and Srivastava, M. S. (1988). Comparison of powers for sphericity tests using both the asymptotic distribution and the bootstrap method. *Commun. Statist. Theory Methods* **17**, 671–690.

Chang, M. N., and Rao, P. V. (1993). Improved estimation of survival functions in the new-better-than-used class. *Technometrics* **35**, 192–203.

Chang, S. I., and Eng, S. S. A. H. (1997). Classification of process variability using neural networks with bootstrap sampling scheme. In *Proceedings of the 6th*

Industrial Engineering Research Conference (G. I. Curry, B. Bidanda, and S. Jagdale, editors), pp. 83–88.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Statist.* **11**, 265–270.

Chao, A., and Huwang, L.-C. (1987). A modified Monte-Carlo technique for confidence limits of system reliability using pass-fail data. *IEEE Trans. Reliab.* **R-36**, 109–112.*

Chao, M.-T., and Lo, S.-H. (1985). A bootstrap method for finite populations. *Sankhya A* **47**, 399–405.*

Chao, M.-T., and Lo, S.-H. (1994). Maximum likelihood summary and the bootstrap method in structured finite populations. *Statist. Sin.* **4**, 389–406.*

Chapman, P., and Hinkley, D. V. (1986). The double bootstrap, pivots and confidence limits. Technical Report 34, Center for Statistical Sciences, University of Texas at Austin.

Chateau, F., and Lebart, L. (1996). Assessing sample variability in the visualization techniques related to principal component analysis: bootstrap and alternative simulation methods. In *COMPSTAT 96 12 Biannual Symposium* (A. Prat, editor), Vol. 12, pp. 205–210. Physica-Verlag, Heidelberg.

Chatfied, C. (1988). *Problem Solving: A Statisticians Guide*. Chapman & Hall, London.

Chatterjee, S. (1984). Variance estimation in factor analysis, an application of bootstrap. *Br. J. Math. Statist. Psychol.* **37**, 252–262.

Chatterjee, S. (1986). Bootstrapping ARMA models. Some simulations. *IEEE Trans. Syst. Man. Cybernet.* **16**, 294–299.

Chatterjee, S., and Chatterjee, S. (1983). Estimation of misclassification probabilities by bootstrap methods. *Commun. Statist. Simul. Comput.* **12**, 645–656.*

Chatterjee, S., and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*. Wiley, New York.*

Chaubey, Y. P. (1993). Book review of *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (by P. H. Westfall, and S. S. Young). *Technometrics* **35**, 450–451.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Am. Statist. Assoc.* **91**, 863–872.

Chen, C., Davis, R. A., Brockwell, P. J., and Bai, Z. D. (1993). Order determination for autoregressive processes using resampling methods. *Statist. Sin.* **3**, 481–500.*

Chen, C., and George, S. I. (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Statist. Med.* **4**, 39–46.*

Chen, C., and Yang, G. L. (1993). Conditional bootstrap procedure for reconstruction of the incubation period of AIDS. *Math. Biosci.* **117**, 253–269.

Chen, H., and Liu, H. K. (1991). Checking the validity of the bootstrap. *Proceedings of the 23rd Symposium on the Interface between Computer Science and Statistics*, pp. 293–296. Springer-Verlag, New York.

Chen, H., and Loh, W. Y. (1991). Consistency of bootstrap for the transformed two sample *t* tests. *Commun. Statist. Theory Methods* **20**, 997–1014.

Chen, H., and Sitter, R. R. (1993). Edgeworth expansions and the bootstrap for stratified sampling without replacement from a finite population. *Can. J. Statist.* **21**, 347–357.

Chen, H., and Tu, D. (1987). Estimating the error rate in discriminant analysis by the delta, jackknife and bootstrap methods. *Chin. J. Appl. Probab. Statist.* **3**, 203–210.

Chen, K., and Lo, S. H. (1996). On bootstrap accuracy with censored data. *Ann. Statist.* **24**, 569–595.

Chen, L. (1995). Testing the mean of skewed distributions. *J. Am. Statist. Assoc.* **90**, 767–772.

Chen, S. X. (1994). Comparing empirical likelihood and bootstrap hypothesis tests. *Multivar. Anal.* **51**, 277–293.

Chen, Z. (1990). A resampling approach for bootstrap hypothesis testing. Unpublished manuscript.

Chen, Z., and Do, K.-A. (1992). Importance resampling for smoothed bootstrap. *J. Statist. Comput. Simul.* **40**, 107–124.

Chen, Z., and Do, K.-A. (1994). The bootstrap methods with saddlepoint approximations and importance resampling. *Statist. Sin.* **4**, 407–421.

Cheng, P. (1993). Weak convergence and bootstrap of bivariate product limit estimation under the bivariate competing risks case (in Chinese). *J. Math. Res. Exp.* **13**, 491–498.

Cheng, P., and Hong, S.-Y. (1993). Bootstrap approximation of parameter estimator in semiparametric regression models (in Chinese). *Sci. Sin.* **23**, 237–251.

Cheng, R. C. H. (1995). Bootstrap methods in computer simulation experiments. In *Proceedings of the 1995 Winter Simulation Conference WSC* '95, pp. 171–177.

Cheng, R. C. H., Holland, W., and Hughes, N. A. (1996). Selection of input models using bootstrap goodness-of-fit. In *Proceedings of the 1996 Winter Simulation Conference*, pp. 199–206.

Chenier, T. C., and Vos, P. W. (1996). Beat Student's t and the bootstrap: Using SAS software to generate conditional confidence intervals for means. *SUGI 21* **1**, 1311–1316.

Chernick, M. R. (1982). The influence function and its application to data validation. *Am. J. Math. Manag. Sci.* **2**, 263–288.*

Chernick, M. R., and Murthy, V. K. (1985). Properties of bootstrap samples. *Am. J. Math. Manag. Sci.* **5**, 161–170.*

Chernick, M. R., Murthy, V. K., and Nealy, C. D. (1985). Applications of bootstrap and other resampling techniques: Evaluation of classifier performance. *Pattern Recogn. Lett.* **3**, 167–178.*

Chernick, M. R., Murthy, V. K., and Nealy, C. D. (1986). Correction note to a pplications of bootstrap and other resampling techniques: Evaluation of classifier performance. *Pattern Recogn. Lett.* **4**, 133–142.*

Chernick, M. R., Murthy, V. K., and Nealy, C. D. (1988a). Estimation of error rate for linear discriminant functions by resampling non-Gaussian populations. *Comput. Math. Applic.* **15**, 29–37.*

Chernick, M. R., Murthy, V. K., and Nealy, C. D. (1988b). Resampling-type error rate estimation for linear discriminant functions: Pearson VII distributions. *Comput. Math. Applic.* **15**, 897–902.*

Chernick, M. R., Daley, D. J., and Littlejohn, R. P. (1988). A time-reversibility relationship between two Markov chains with exponential stationary distributions. *J. Appl. Probab.* **25**, 418–422.*

Chernick, M. R., Downing, D. J., and Pike, D. H. (1982). Detecting outliers in time series data. *J. Am. Statist. Assoc.* **77**, 743–747.*

Chin, L., Haughton, D., and Aczel, A. (1996). Analysis of student evaluation of teaching scores using permutation methods. *J. Comput. Higher Educ.* **8**, 69–84.

Cho, K., Meer, P., and Cabrera, J. (1997). Performance assessment through bootstrap. *IEEE Trans. Pattern Recog. Mach. Intell.* **19**, 1185–1198.

Choi, K. C., Nam, K. H., and Park, D. H. (1996). Estimation of capability index based on bootstrap method. *Microelectron. Reliab.* **36**, 256–259.*

Choi, S. C. (1986). Discrimination and classification: Overview. *Comput. Math. Applic.* **12A**, 173–177.*

Christofferson, J. (1997). Frequency domain resampling of time series data. Ph.D. thesis. Acta Universitatis Agriculturae Sueviae, Silvestria.

Chu, P.-S., and Wang, J. (1998). Interval variability of tropical cyclone incidences in the vicinity of Hawaii using statistical resampling techniques. *Conference on Probability and Statistics in the Atmosphere Sciences*, **14**, 96–97.

Chung, C.-J. F. (1989). Confidence bands for percentile residual lifetime under random censorship model. *J. Multivar. Anal.* **29**, 94–126.

Chung, K. H., and Lee, S. M. S. (1996). Optimal bootstrap sample size in construction of percentile confidence intervals. *Research Report 127*. Department of Statistics, University of Hong Kong.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.*

Ciarlini, P. (1997). Bootstrap algorithms and applications. In *Advanced Mathematical Tools in Metrology III* (P. Ciarlini, M. G. Cox, F. Paveses, and D. Richter, editors), pp. 171–177. World Scientific, Singapore.

Ciarlini, P., Gigli, A., Regoliosi, G., Moiraghi, L., and Montefusco, A. (1996). Monitoring an industrial process: bootstrap estimation of accuracy of quality parameters. In *Advanced Mathematical Tools in Metrology II* (P. Ciarlini, M. G. Cox, F. Paveses, and D. Richter, editors), pp. 123–129. World Scientific, Singapore.

Cirincione, C., and Gurrieri, G. A. (1997). Research methodology. Computer-intensive methods in the social sciences. *Soc. Sci. Comput. Res.* **15**, 83–97.

Clayton, H. R. (1994). Bootstrap approaches: some potential problems. *1994 Proceedings Decision Science Institute*, Vol. 2, pp. 1394–1396.

Clayton. H. R. (1996). Developing a robust parametric bootstrap bound for evaluating audit samples. *Proceedings of the Annual Meeting of the Decision Sciences Institute*, Vol. 2, pp. 1107–1109. Decision Sciences Institute, Atlanta.

Cleveland, W. S., and McGill, R. (1983). A color-caused optical illusion on a statistical graph. *Am. Statist.* **37**, 101–105.

Cliff, A. D., and Ord, J. K. (1973). *Spatial Autocorrelation*. Pion, London.*

Cliff, A. D., and Ord, J. K. (1981). *Spatial Processes: Models and Applications*. Pion, London.*

Coakley, K. J. (1991). Bootstrap analysis of asymmetry statistics for polarized beam studies. *Proceedings of the 23rd Symposium on the Interface Between Computer Science and Statistics*, pp. 301–304, Springer-Verlag, New York.

Coakley, K. J. (1994). Area sampling scheme for improving maximum likelihood reconstructions for positive emission tomography images. *Proceedings of Medical Imaging 1994, The International Society for Optical Engineering (SPIE)*, Vol. 2167, pp. 271–280.

Coakley, K. J. (1996). Bootstrap method for nonlinear filtering of EM-ML reconstructions of PET images. *Int. J. Imaging Syst. Technol.* **7**, 54–61.*

Cochran, W. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York.*

Cohen, A. (1986). Comparing variances of correlated variables. *Psychometrika* **51**, 379–391.

Cohen, A., Lo, S.-H., and Singh, K. (1985). Estimating a quantile of a symmetric distribution. *Ann. Statist.* **13**, 1114–1128.

Cole, M. J., and McDonald, J. W. (1989). Bootstrap goodness-of-link testing in generalized linear models. In *Statistical Modelling: Proceedings of GLIM 89 and the 4th International Workshop on Statistical Modelling* (A. Decarli, B. J. Francis, R. Gilchrist, and G. L. H. Seeher, editors), *Lecture Notes in Statistics*, Vol. 57. pp. 84–94. Springer-Verlag, Berlin.

Collings, B. J., and Hamilton, M. A. (1988). Estimating the power of the two-sample Wilcoxon test for location shift. *Biometrics* **44**, 847–860.

Constantine, K., Karson, M. J., and Tse, S.-K. (1990). Confidence interval estimation of $P(Y < X)$ in the gamma case. *Commun. Statist. Simul. Comput.* **19**, 225–244.

Conti, P. L. (1993). Book Review of *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors). *Metron* **51**, 213–217.

Cook, R. D., and Weisberg, S. (1990). Confidence curves in nonlinear regression. *J. Am. Statist. Assoc.* **85**, 544–551.

Cook, R. D., and Weisberg, S. (1994). Transforming a response variable for linearity. *Biometrika* **81**, 731–737.*

Cowling, A., Hall, P., and Phillips, M. J. (1996). Bootstrap confidence regions for the intensity of a Poisson point process. *J. Am. Statist. Assoc.* **91**, 1516–1524.*

Cover, K. A., Unny, T. E. (1986). Applicatiion of computer intensive statistics to parameter uncertainty in streamflow. *Water Res. B* **22**, 495–507.

Cox, D. R. (1972). Regression models and life tables. *J. R. Statist. Soc. B* **34**, 187–202.*

Cox, D. R., and Reid, N. (1987a). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B* **49**, 1–39.*

Cox, D. R., and Reid, N. (1987b). Approximations to noncentral distributions. *Can. J. Statist.* **15**, 105–114.*

Crawford, S. (1989). Extensions to the CART algorithm. *Int. J. Mach. Studies* **31**, 197–217.

Cressie, N. (1991). *Statistics for Spatial Data*. Wiley, New York.*

Cressie, N. (1993). *Statistics for Spatial Data*. Revised Edition. Wiley, New York.*

Crivelli, A., Firimguetti, L., Montano, R., and Munoz, M. (1995). Confidence intervals in ridge regression by bootstrappimg the dependent variable : a simulation study. *Commun. Statist. Simul. Comput.* **24**, 631–652.

Croci, M. (1994). Some semi-parametric bootstrap applications in unfavourable circumstances (in Italian). *Proc. Ital. Statist. Soc.* **2**, 449–456.

Crone, L. J., and Crosby, D. S. (1995). Statistical applications of a metric on subspaces to satellite meteorology. *Technometrics* **37**, 324–328.

Crosilla, F., and Pillirone, G. (1994). Testing the data quality of G.I.S. data base by bootstrap methods and other nonparametric statistics. *ISPRS Commission III Symposium on Spatial Information from Photogrammetry and Computer Vision, Munich Germany, Sept. 5–9, 1994. SPIE Proce.* Vol. 2357, pp. 158–164.

Crosilla, F., and Pillirone, G. (1995). Non-parametric statistics and bootstrap methods for testing data quality of a geographic information system. In *Geodetic Theory Today: Symposium of the International Association of Geodesy* (F. Sansao, editor), Vol. 114, pp. 214–223.

Crowder, M. J., and Hand, D. J. (1990). *Analysis of Repeated Measures*. Chapman & Hall, London.

Crowley, P. H. (1992). Resampling methods for computer-intensive data analysis in ecology and evolution. *Annu. Rev. Ecol. Syst.* **23**, 405–407.

Csorgo, M. (1983). *Quantile Processes with Statistical Applications*. SIAM, Philadelphia.*

Csorgo, M., Csorgo, S., and Horvath, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes*. Springer-Verlag, New York.*

Csorgo, M., and Zitikis, R. (1996). Mean residual life processes. *Ann. Statist.* **24**, 1717–1739.

Csorgo, S., and Mason, D. M. (1989). Bootstrapping empirical functions. *Ann. Statist.* **17**, 1447–1471.

Cuesta-Albertos, J. A., Gordaliza, A., and Matran, C. (1997). Trimmed *k*-means: An attempt to robustify quantizers. *Ann. Statist.* **25**, 553–576.

Cuevas, A., and Romo, J. (1994). Continuity and differentiability of statistical operators: some applications to the bootstrap. In *Proceedings of the 3rd World Conference of the Bernoulli Society and the 57th Annual Meeting of the Institute of Mathematical Statistics, Chapel Hill*.

Czado, C. (2000). Multivariate regression analysis of panel data with binary outcomes applied to unemployment data. *Statist. Papers* **41**, 281–304.

Dabrowski, D. M. (1989). Kaplan-Meier estimate on the plane: Weak convergence, LIL, and the bootstrap. *J. Multivar. Anal.* **29**, 308–325.

Daggett, R. S., and Freedman, D. A. (1985). Econometrics and the law: A case study in the proof of antitrust damages. *Proc. Berk. Symp.* **VI**, 123–172.*

Dahlhaus, R., and Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.* **24**, 1914–1933.*

Dalal, S. R., Fowlkes, E. B., and Hoadley, B. (1989). Risk analysis of the space shuttle pre-*Challenger* prediction of failure. *J. Am. Statist. Assoc.* **84**, 945–957.

Daley, D. J., and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York.*

Dalgleish, L. I. (1995). Software review: Bootstrapping and jackknifing with BOJA. *Statist. Comput.* **5**, 165–174.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–650.*

Daniels, H. E., and Young, G. A. (1991). Saddlepoint approximations for the studentized mean, with an application to the bootstrap. *Biometrika* **78**, 169–179.*

Das, S., and Sen, P. K. (1995). Simultaneous spike-trains and stochastic dependence. *Sankhya B* **57**, 32–47.

Das Peddada, S., and Chang, T. (1992). Bootstrap confidence region estimation of motion of rigid bodies. *J. Am. Statist. Assoc.* **91**, 231–241.*

Datta, S. (1992). A note on the continuous Edgeworth expansions and bootstrap. *Sankhya A* **54**, 171–182.

Datta, S. (1994). On a modified bootstrap for certain asymptotically nonnormal statistics. *Statist. Probab. Lett.* **24**, 91.

Datta, S. (1995). Limit theory for explosive and partially explosive autoregression. *Stoch. Proc. Appl.* **57**, 285–304.*

Datta, S. (1996). On asymptotic properties of bootstrap for AR(1) processes. *J. Statist. Plann. Inf.* **53**, 361–374.*

Datta, S., and McCormick, W. P. (1992). Bootstrap for finite state Markov chain based on i. i. d. resampling. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 77–97. Wiley, New York.*

Datta, S., and McCormick, W. P. (1995a). Bootstrap inference for a first-order autoregression with positive innovations. *J. Am. Statist. Assoc.* **90**, 1289–1300.*

Datta, S., and McCormick, W. P. (1995b). Some continuous Edgeworth expansions for Markov chains with applications to bootstrap. *J. Mult. Anal.* **52**, 83–106.

Datta, S., and Sriram (1997). A modified bootstrap for autoregression without stationarity. *J. Statist. Plann. Inf.* **59**, 19–30.*

Daudin, J. J., Duby, C., and Trecourt, P. (1988). Stability of principal component analysis studied by bootstrap method. *Statistics* **19**, 241–258.

David, H. A. (1981). *Order Statistics*. Wiley, New York.*

Davis, C. E., and Steinberg, S. M. (1986). Quantile estimation. In *Encyclopedia of Statistical Sciences*, Vol. 7, pp. 408–412. Wiley, New York.

Davis, R. B. (1995). Resampling a good tool, no panacea. *M. D. Comput.* **12**, 89–91.

Davison, A. C. (1993). Book Review of *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors). *J. R. Statist. Soc. A* **156**, 133.

Davison, A. C., and Hall, P. (1992). On the bias and variability of bootstrap and cross-validation estimates of error rate in discriminant analysis. *Biometrika* **79**, 279–284.*

Davison, A. C., and Hall, P. (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Aust. J. Statist.* **35**, 215–224.*

Davison, A. C., and Hinkley, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417–431.*

Davison, A. C., and Hinkley, D. V. (1992). Computer-intensive statistical methods. *Proceedings of the 10th Symposium on Computational Statistics* **2**, 51–62.

Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.*

Davison, A. C., Hinkley, D. V., and Schechtman, E. (1986). Efficient bootstrap simulation. *Biometrika* **73**, 555–566.*

Davison, A. C., Hinkley, D. V., and Worton, B. J. (1992). Bootstrap likelihoods. *Biometrika* **79**, 113–130.*

Davison, A. C., Hinkley, D. V., and Worton, B. J. (1995). Accurate and efficient construction of bootstrap likelihoods. *Statist. Comput.* **5**, 257–264.

Day, N. E. (1969). Estimating the components of a mixture of two normal distributions. *Biometrika* **56**, 463–470.*

DeAngelis, D., and Young, G. A. (1992). Smoothing the bootstrap. *Int. Statist. Rev.* **60**, 45–56.*

DeAngelis, D., and Young, G. A. (1998). Bootstrap method. In *Encyclopedia of Biostatistics* (P. Armitage, and T. Colton, editors), Vol. 1, pp. 426–433. Wiley, New York.*

DeAngelis, D., Hall, P., and Young, G. A. (1993a). Analytical and bootstrap approximations to estimator distributions in L^1 regression. *J. Am. Statist. Assoc.* **88**, 1310–1316.*

DeAngelis, D., Hall, P., and Young, G. A. (1993b). A note on coverage error of bootstrap confidence intervals for quantiles. *Math. Proc. Cambridge Philos. Soc.* **114**, 517–531.

Deaton, M. L. (1988). Simulation models (validation of). In *Encyclopedia of Statistical Sciences*, Vol. 8, pp. 481–484. Wiley, New York.

DeBeer, C. F., and Swanepoel, J. W. H. (1989). A modified Durbin-Watson test for serial correlation in multiple regression under non-normality using bootstrap. *J. Statist. Comput. Simul.* **33**, 75–82.

DeBeer, C. F., and Swanepoel, J. W. H. (1993). A modified bootstrap estimator for the mean of an asymmetric distribution. *Can. J. Statist.* **21**, 79–87.

Deheuvels, P., Mason, D. M., and Shorack, G. R. (1993). Some results on the influence of extremes on the bootstrap. *Ann. Inst. Henri Poincare* **29**, 83–103.

de Jongh, P. J., and De Wet, T. (1985). A Monte Carlo comparison of regression trimmed means. *Commun. Statist. Simul.Comput.* **14**, 2457–2472.

de Jongh, P. J., and De Wet, T. (1986). Confidence intervals for regression parameters based on trimmed means. *S. Afr. Statist. J.* **20**, 137–164.

Delaney, N. J., and Chatterjee, S. (1986). Use of bootstrap and cross-valiidation in ridge regression. *J. Bus. Econ. Statist.* **4**, 255–262.

de la Pena, V., Gini, E., and Alemayehu, D. (1993). Bootstrap goodness-of-fit tests based on the empirical characteristic function. *Proceedings of the 25th Symposium on the Interface of Computer Science and Statistics*, Vol. 25, pp. 228–233.

Delicado, P., and Del Rio, M. (1994). Bootstrapping the general linear hypothesis test. *Comput. Statist. Data Anal.* **18**, 305–316.

Dempster, A. P. Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B* **39**, 1–38.*

Dennis, B., and Taper, M. L. (1994). Density dependence in time series observations of natural populations: Estimation and testing. *Ecol. Monogr*. **64**, 205–224.

Depuy. K. M., Hobbs, J. R., Moore, A. H., and Johnson, J. W., Jr. (1982). Accuracy of univariate, bivariate and a "modified double Monte Carlo" technique for finding lower confidence limits of system reliability. *IEEE Trans. Reliab* **R-31**, 474–477.

Desgagne, A., Castilloux, A.-M., Angers, J.-F., and Lelorier, J. (1998). The use of the bootstrap statistical method for the pharmacoeconomic cost analysis of skewed data. *Pharmacoeconomics* **13**, 487–497.

Dette, H., and Munk, A. (1998). Validation of linear regression models. *Ann. Statist*. **26**, 778–800.

Devorye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.*

Devorye, L., and Gyorfi, L. (1985). *Nonparametric Density Estimation*. Wiley, New York.*

de Wet, T., and Van Wyk, J. W. J. (1986). Bootstrap confidence intervals for regression coefficients when the residuals are dependent. *J. Statist. Comput. Simul*. **23**, 317–327.

Diaconis, P. (1985). Theories of data analysis from magical thinking through classical statistics. In *Exploring Data Tables, Trends, and Shapes* (D. C. Hoaglin, F. Mosteller, and J. W. Tukey, editors), pp. 1–36. Wiley, New York.

Diaconis, P., and Efron, B. (1983). Computer-intensive methods in statistics. *Sci. Am*. **248**, 116–130.*

Diaconis, P., and Holmes, S. (1994). Gray codes for randomization procedures. *Statist. Comput*. **4**, 287–302.*

DiCiccio, T. J., and Efron, B. (1990). Better approximate confidence intervals in exponential families,. Technical Report 345, Department of Statistics, Stanford University.*

DiCiccio, T. J., and Efron, B. (1992). More Accurate confidence intervals in exponential families. *Biometrika* **79**, 231–245.*

DiCiccio, T. J., and Efron, B. (1996). Bootstrap confidence intervals (with discussion). *Statist. Sci*. **11**, 189–228.*

DiCiccio, T. J., Hall, P., and Romano, J. P. (1989). Comparison of parametric and empirical likelihood functions. *Biometrika*, **76**, 465–476.

DiCiccio, T. J., Hall, P., and Romano, J. P. (1991). Empirical likelihood is Bartlett correctable. *Ann. Statist*. **19**, 1053–1061.

DiCiccio, T. J., Martin, M. A., and Young, G. A. (1992a). Fast and accurate approximate double bootstrap distribution functions using saddlepoint methods, *Biometrika* **79**, 285–295.*

DiCiccio, T. J., Martin, M. A., and Young, G. A. (1992b). Analytic approximations for iterated bootstrap confidence intervals. *Statist. Comput*. **2**, 161–171.

DiCiccio, T. J., Martin, M. A., and Young, G. A. (1994). Analytic approximations to bootstrap distribution functions using saddlepoint methods. *Statist. Sin*. **4**, 281–295.

DiCiccio, T. J., and Romano, J. P. (1988). A review of bootstrap confidence intervals (with discussion). *J. R. Statist. Soc. B* **50**, 338–370. Correction. *J. R. Statist. Soc. B* **51**, 470.*

DiCiccio, T. J., and Romano, J. P. (1989a). The automatic percentile method: accurate confidence limits in parametric models. *Can. J. Statist*. **17**, 155–169.

DiCiccio, T. J., and Romano, J. P. (1989b). On adjustments based on the signed root of the empirical likelihood ratio statistic. *Biometrika* **76**, 447–456.

DiCiccio, T. J., and Romano, J. P. (1990). Nonparametric confidence limits by resampling methods and least favorable families. *Int. Statist. Rev*. **58**, 59–76.

DiCiccio, T. J., and Tibshirani, R. (1987). Bootstrap confidence intervals and bootstrap approximations. *J. Am. Statist. Assoc*. **82**, 163–170.

Diebolt, J., and Ip, E. H. S. (1995). Stochastic EM: Method and Application. In *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors), pp. 259–273. Chapman & Hall, London.

Dielman, T. E., and Pfaffenberger, R. C. (1988). Bootstrapping in least absolute value regression: an application to hypothesis testing. *Commun. Statist. Simul. Comput*. **17**, 843–856.*

Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, New York.*

Diggle, P. J., Lange, N., and Benes, F. M. (1991). Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *J. Am. Statist. Assoc*. **78**, 618–625.*

Dijkstra, D. A., and Veldkamp, J. H. (1988). Data-driven selection of regressors and the bootstrap. In *On Model Uncertainty and Its Statistical Implications*. (T. K. Dijkstra, editor), *Lecture Notes in Economics and Mathematical Systems*, Vol. 307. Springer-Verlag, Berlin.

Dikta, G. (1990). Bootstrap approximation of nearest neighbor regression function estimates. *J. Multivar. Anal*. **32**, 213–229.*

Dikta, G., and Ghorai, J. K. (1990). Bootstrap approximation with censored data under the proportional hazard model. Commun. *Statist. Theory Methods* **19**, 573–581.

Dirschedl, P., and Grohmann, R. (1992). Exploring heterogeneous risk structure: Comparison of bootstrapped model selection and nonparametric classification Technique. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 189–195. Springer-Verlag, Berlin.

Dixon, P. M. (1993). The bootstrap and the jackknife: describing the precision of ecological studies. In *Design and Analysis of Ecological Experiments* (S. M. Scheiner and J. Gurevitch, editors), pp. 290–318. Chapman & Hall, New York.

Dixon, P. M., Weiner, J., Mitchell-Olds, T., and Woodley, R. (1987). Bootstrapping the Gini coefficient of inequality. *Ecology* **68**, 1548–1551.

Do, K.-A. (1992). A simulation study of balanced and antithetic bootstrap resampling methods. *J. Statist Comput. Simul*. **40**, 153–156.*

Do, K.-A. (1996). A numerical study of the matched-block bootstrap for dependent data. In *20th Symposium on the Interface Between Computer Science and Statistics* (M. M. Meyer, and J. L. Rosenberger, editors) Vol. 27, pp. 425–429.

Do, K.-A., and Hall, P. (1991a). On importance resampling for the bootstrap. *Biometrika* **78**, 161–167.*

Do, K.-A., and Hall, P. (1991b). Quasi-random resampling for the bootstrap. *Statist. Comput*. **1**, 13–22.*

Do, K.-A., and Hall, P. (1992). Distribution estimation using concomitants of order statistics with applications to Monte Carlo simulation for the bootstrap. *J. R. Statist. Soc. B* **54**, 595–607.

Dohman, B. (1990). Confidence intervals for small sample sizes: Bootstrap vs. standard methods (in German), Diplom. Thesis. University of Siegen.

Donegani, M. (1992), A bootstrap adaptive test for two-way analysis of variance. *Biomed. J*. **34**, 141–146.

Dopazo, J. (1994). Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. *J. Mol. Evol*. **38**, 300–304.

Dorfman, D. D., Berbaum, K. S., and Lenth, R. V. (1995). Multireader, multicase, receiver operating characteristic methodology: A bootstrap analysis. *Acad. Radiol*. **2**, 626–633,

Doss, H., and Chiang, Y. C. (1994). Choosing the resampling scheme when bootstrapping: A case study in reliability. *J. Am. Statist, Assoc*. **89**, 298–308.

Doss, H., and Gill, R. D. (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *J. Am. Statist. Assoc*. **87**, 869–877.

Douglas, S. M. (1987). Improving the estimation of a switching regressions model: an analysis of problems and improvementd using the bootstrap. Ph.D. dissertation. University of North Carolina.

Draper, N. R., and Smith, H. (1981). *Applied Regression Analysis*, 2nd ed. Wiley, New York.*

Draper, N. R., and Smith, H. (1998). *Applied Regression Analysis*, 3rd ed. Wiley, New York.*

Droge, B. (1987). A note on estimating the MSEP in nonlinear regression. *Statistics* **18**, 499–520.

Duan, N. (1983). Smearing estimate a nonparametric retransformation method. *J. Am. Statist. Assoc*. **78**, 605–610.*

Ducharme, G. R., and Jhun, M. (1986). A note on the bootstrap procedure for testing linear hypotheses. *Statistics* **17**, 527–531.

Ducharme, G. R., Jhun, M., Romano, J., and Truong, K. N. (1985). Bootstrap confidence cones for directional data. *Biometrika* **72**, 637–645.

Duda, R. O., and Hart, P. E. (1973). *Pattern Recognition and Scene Analysis*. Wiley, New York.*

Dudewicz, E. J. (1992). The generalized bootstrap. In *Bootstrapping and Related Techniques, Proceedings, Trier, FRG*. (K.-H. Jockel, G, Rothe, and W. Sendler, editors), *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 31–37. Springer-Verlag, Berlin.*

Dudewicz, E. J. (editor) (1996). *Modern Digital Simulation Methodology, II: Univariate and Bivariate Distribution Fitting, Bootstrap Methods & Application*. American Sciences Press, Columbus, OH.

Dudewicz, E. J., and Mishra, S. N. (1988). *Modern Mathematical Statistics*. Wiley, New York.*

Dudley, R. M. (1978). Central limit theorem for empirical measures. *Ann. Probab*. **6**, 899–929.

DuMouchel, W., and Waternaux, C. (1983). Some new dichotomous regression methods. In *Recent Advances in Statistics* (M. H. Rizvi, J. Rustagi, and D. Siegmund, editors), pp. 529–555. Academic Press, New York.

Dumbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist*. **19**, 1471–1495.

Dutendas, D., Moreau, L., Ghorbel, F., and Allioux, P. M. (1995). Unsupervised Bayesian segmentation with bootstrap sampling application to eye fundus image coding. In *IEEE Nuclear Science Symposium and Medical Imaging Conference*, **4**, 1794–1796.

Eakin, B. K., McMillen, D. P., and Buono, M. J. (1990). Constructing confidence intervals using the bootstrap: An application to multi-product cost function. *Rev. Econ. Statist*. **72**, 339–344.

Eaton, M. L., and Tyler, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist*. **19**, 260–271.

Ecker, M. D., and Heltsche, J. F. (1994). Geostatistical estimates of scallop abundance: In *Case Studies in Biometry*. (N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest, and J. Greenhouse, editors), pp. 107–124, Wiley, New York.

Eckert, R. S., Carroll, R. J., and Wang, W. (1997). Transformations to additivity in measurement error models. *Biometrics* **53**, 262–272.

Eddy, W. F., and Gentle, J. E. (1985). Statistical computing: what's past is prologue. In *A Celebration of Statistics. The ISI Centenary Volume* (A. C. Atkinson and S. E. Feinberg, editors), pp. 233–249. Springer-Verlag, New York.

Edgington, E. S. (1986). Randomization tests. In *Encyclopedia of Statistical Sciences*, Vol. 7, pp. 530–538, Wiley, New York.

Edgington, E. S. (1980). *Randomization Tests*, Marcel Dekker, New York.*

Edgington, E. S. (1987). *Randomization Tests*, 2nd ed. Marcel Dekker, New York.*

Edgington, E. S. (1995). *Randomization Tests*, 3rd ed. Marcel Dekker, New York.*

Edler, L., Groger, P., and Thielmann, H. W. (1985). Computational statistics for cell survival curves II: evaluation of colony-forming ability of a group of cell strains by the APL function CFAGROUP. *Comp. Biomed*. **21**, 47–54.

Efron, B. (1978). Controversies in the foundations of statistics. *Am. Math. Month*. **85**, 231–246.*

Efron, B. (1979a). Bootstrap methods; another look at the jackknife. *Ann. Statist*. **7**, 1–26.*

Efron, B. (1979b). Computers and the theory of statistics: thinking the unthinkable. *SIAM Rev*. **21**, 460–480.*

Efron, B. (1981a). Censored data and the bootstrap. *J. Am. Statist. Assoc*. **76**, 312–319.*

Efron, B. (1981b). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika* **68**, 589–599.

Efron, B. (1981c). Nonparametric standard errors and confidence intervals (with discussion). *Can, J, Statist*. **9**, 139–172.

Efron, B. (1982a). *The Jackknife, the Bootstrap, and Other Resampling Plans*. SIAM, Philadelphia.*

Efron, B. (1982b). Computer-intensive methods in statistics. In *Some Recent Advances in Statistics*. (J. T. de Oliveira and B. Epstein, editors), pp. 173–181. Academic Press, London.*

Efron, B. (1983). Estimating the error rate of a prediction rule: improvements on cross-validation. *J. Am. Statist. Assoc*. **78**, 316–331.*

Efron, B. (1984). Comparing non-nested linear models. *J Am. Statist. Assoc*. **79**, 791–803.

Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika*, **72**, 45–58.

Efron, B. (1986). How biased is the apparent error rate of a prediction rule? *J. Am. Statist. Assoc*. **81**, 461–470.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Am. Statist. Assoc*. **82**, 171–200.*

Efron, B. (1988a). Three examples of computer-intensive statistical inference. *Sankhya A* **50**, 338–362.*

Efron, B. (1988b). Computer-intensive methods in statistical regression. *SIAM Rev*. **30**, 421–449.*

Efron, B. (1988c). Bootstrap confidence intervals. Good or bad? *Psychol. Bull*. **104**, 293–296.*

Efron, B. (1990). More efficient bootstrap computations. *J. Am. Statist. Assoc*. **85**, 79–89.*

Efron, B. (1992a). Regression percentiles using asymmetric squared error loss. *Statist. Sin*. **1**, 93–125.*

Efron, B. (1992b). Six questions raised by the bootstrap. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 99–126. Wiley, New York.*

Efron, B. (1992c). Jackknife-after-bootstrap standard errors and influence functions (with discussion). *J. R. Statist. Soc. B* **54**, 83–127.*

Efron, B. (1993). Statistics in the 21st century. *Statist. Comput*. **3**, 188–190.

Efron, B. (1994). Missing data, imputation, and the bootstrap (with discussion). *J. Am. Statist. Assoc*. **89**, 463–479.*

Efron, B. (1996). Empirical Bayes methods for combining likelihoods (with discussion). *J. Am. Statist. Assoc*. **91**, 538–565.

Efron, B., and Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. *J. Am. Statist. Assoc*. **86**, 9–26.

Efron, B., and Gong, G. (1981). Statistical thinking and the computer. In Proceedings Computer Science and Statistics (W. F. Eddy, editor), Vol. 13, pp. 3–7. Springer-Verlag, New York.

Efron, B., and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife and cross-validation. *Am. Statist*. **37**, 36–48.*

Efron, B., Halloran, M. E., and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci*. **93**, 13429–13434.

Efron, B., and LePage, R. (1992). Introduction to bootstrap. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 3–10, Wiley, New York.*

Efron, B., and Stein, C. (1981). The jackknife estimate of variance, *Ann.Statist*. **9**, 586–596.

Efron, B., and Tibshirani, R. (1985). The bootstrap method for assessing statistical accuracy. *Behaviormetrika* **17**, 1–35.*

Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors: Confidence intervals and other measures of statistical accuracy. *Statist. Sci*. **1**, 54–77.*

Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.*

Efron, B., and Tibshirani, R. (1996a). Computer-intensive statistical methods. In *Advances in Biometry* (P. Armitage and H. A. David, editors), pp. 131–147. Wiley, New York.*

Efron, B., and Tibshirani, R. (1996b). The problem of regions. Stanford University Technical Report No. 192.*

Efron, B., and Tibshirani, R. (1997a). Improvements on cross-validation: The .632+ bootstrap methods. *J. Am. Statist. Assoc*. **92**, 548–560.*

Efron, B., and Tibshirani, R. (1997b). Computer-intensive statistical methods. In *Encyclopedia of Statistical Sciences*, Update Volume 1 (S. Kotz, C. B. Read, and D. L. Banks, editors), pp. 139–148. Wiley, New York.*

Efron, B., and Tibshirani, R. (1998). The problem of regions. *Ann. Statist*. **26**, 1687–1718.

El-Sayed, S. M., Jones, P. W., and Ashour, S. K. (1991). Bootstrap censored data analysis of fractured femur patients under the mixed exponential model using maximum likelihood. In *26th Conference on Statistics, Computer Science and Operations Research: Mathematical Statistics*, Vol. 1, pp. 39–54. Cairo University.

English, J. R., and Taylor, G. D. (1990). Process capability analysis—A robustness study. Master of Science, Department of Industrial Engineering, University of Arkansas at Fayetteville.*

Eriksson, B. (1983). On the construction of confidence limits for the regression coefficients when the residuals are dependent. *J. Statist. Comput. Simul*. **17**, 297–309.

Eubank, R. L. (1986). Quantiles. *Encyclopedia of Statistical Sciences*, Vol. 7, pp. 424–432. Wiley, New York.

Eynon, B., and Switzer, P. (1983). The variability of rainfall acidity. *Can. J. Statist*. **11**, 11–24.*

Fabiani, M., Gratton, G., Corballis, P. M., Cheng, J., and Friedman, D. (1998). Bootstrap assessment of the reliability of maxima in surface maps of brain activity of individual subjects derived with electrophysiological and optical methods. *Behav. Res. Methods Instrum. Comput*. **30**, 78–86.

Falck, W., Bjornstad, O. N., and Stenseth, N. C. (1995). Bootstrap estimated uncertainty of the dominant lyapunov exponent for Holarctic mocrotine rodents. *Proc. R. Soc. London B* **261**, 159–165.

Falk, M. (1986a). On the estimation of the quantile density function. *Statist. Probab. Lett*. **4**, 69–73.

Falk, M. (1986b). On the accuracy of the bootstrap approximation of the joint distribution of sample quantiles. *Commun. Statist. Theory Methods* **15**, 2867–2876.

Falk, M. (1988). Weak convergence of the bootstrap process for large quantiles. *Statist. Dec*. **6**, 385–396.

Falk, M. (1990). Weak convergence of the maximum error of the bootstrap quantile estimate. *Statist. Probab. Lett*. **10**, 301–305.

Falk, M. (1991). A note on the inverse bootstrap process for large quantiles. *Stoch. Proc.Their Appl*. **38**, 359–363.

Falk, M. (1992a). Bootstrapping the sample quantile: a survey. In *Bootstrapping and Related Techniques Proceedings, Trier, FRG* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 165–172. Springer-Verlag, Berlin.*

Falk, M. (1992b). Bootstrap optimal bandwidth selection for kernel density estimates. *J. Statist. Plann. Inf*. **30**, 13–32.*

Falk, M., and Kaufmann, E. (1991). Coverage probabilities of bootstrap confidence intervals for quantile estimates. *Ann. Statist*. **19**, 485–495.*

Falk, M., and Reiss, R.-D. (1989a). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *Ann. Probab*. **17**, 362–371.

Falk, M., and Reiss, R.-D. (1989b). Bootstrapping the distance between smooth bootstrap and sample quantile distribution. *Probab. Theory Relat. Fields* **82**, 177–186.

Falk, M., and Reiss, R.-D. (1989c). Statistical inference of conditional curves. Poisson process approach. Preprint 231, University of Seigen.

Falk, M., and Reiss, R.-D. (1992). Bottstrapping conditional curves. In *Bootstrapping and Related Techniques. Proceedings, Tier, FRG* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 173–180. Springer-Verlag, Berlin.

Fan, J., and Lin, J.-T. (1998). Test of significance when data are curves. *J. Am. Statist. Assoc*. **93**, 1007–1021.

Fan, T.-H., and Hung, W.-L. (1997). Balanced resampling for bootstrapping finite Markov chains. *Commun. Statist. Simul. Comput*. **26**, 1465–1475.*

Fang, K. T., and Wang, Y. (1994). *Number-Theoretic Methods in Statistics*. Chapman & Hall, London.*

Faraway, J. J. (1990). Bootstrap selection of bandwidth and confidence bands for nonparametric regression. *J. Statist. Comput. Simul*. **37**, 37–44.

Faraway, J. J. (1992). On the cost of data analysis. *J. Comput. Graphical Statist*. **1**, 213–229.

Faraway, J. J., and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *J. Am. Statist. Assoc*. **85**, 1119–1122.*

Farewell, V. (1985). Nonparametric estimation of standard errors. In *Encyclopedia of Statistical Sciences*, Vol. 6, (S. Kotz, N. L. Johnson, and C. B. Read, editors), pp. 328–331, Wiley, New York.

Farrell, P. J., MacGibbon, B., and Tomberlin, T. J. (1997). Empirical Bayes small-area estimation using logistic regression models and summary statistics. *J. Bus. Econ. Statist*. **15**, 101–110.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. **2**, 2nd ed., Wiley, New York.*

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.*

Felsenstein, J. (1992). Estimating effective population size from samples of sequences: A bootstrap Monte Carlo integration method. *Genet. Res. Cambridge* **60**, 209–220.*

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theorectic Approach*. Academic Press, New York.*

Fernholtz, L. T. (1983). *von Mises Calculus for Statistical Functionals*. Lecture Notes in Statistics, Vol. 19. Springer-Verlag, New York.*

Ferreira, F. P., Stangenhaus, G., and Narula, S. C. (1993). Bootstrap confidence intervals for the minimum sum of absolute errors regression. *J. Statist. Comput. Simul*. **48**, 127–133.

Ferretti, N., and Romo, J. (1996). Unit root bootstrap tests for AR(1) models. *Biometrika* **83**, 849–860.

Field, C., and Ronchetti, F. (1990). *Small Sample Asymptotics. Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Vol. 13. Institute of Mathematical Statistics, Hayward.*

Fiellin, D. A. and Feinstein, A. R. (1998). Bootstraps and jackknives: new computer-intensive statistical tools that require no mathematical theories. *J. Invest. Med*. **46**, 22–26.*

Findley, D. F. (1986). Om bootstrap estimates of forecast mean square errors for autoregressive processes. *Proc. Comput. Sci. Statist*. **17**, 11–17.*

Firth, D., Glosup, J., and Hinkley, D. V. (1991). Model checking with nonparametric curves. *Biometrika* **78**, 245–252.

Fisher, G., and Sim, A. B. (1995). Some finite sample theory for bootstrap regression estimates. *J. Statist. Plann. Inf*. **43**, 289–300.

Fisher, N. I., and Hall, P. (1989). Bootstrap confidence regions for directional data. *J. Am. Statist. Assoc* **84**, 996–1002.*

Fisher, N. I., and Hall, P. (1990). On bootstrap hypothesis testing. *Aust. J. Statist*. **32**, 177–190.*

Fisher, N. I., and Hall, P. (1991). Bootstrap algorithms for small samples. *J. Statist. Plann. Inf*. **27**, 157–169.

Fisher, N. I., and Hall, P. (1992). Bootstrap methods for directional data. In *The Art of Statistical Science* (K. V. Mardia, editor), pp. 47–63, Wiley, New York.

Fisher, N. I., Hall, P., Jing, B.-Y., and Wood, A. T. A. (1996). Improved pivotal methods for constructing confidence regions with directional data. *J. Am. Statist. Assoc*. **91**, 1062–1070.

Fisher, N. I., Lewis, T., and Embleton, B. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge University Press, Cambridge.*

Fitzmaurice, G. M., Laird, N. M., and Zahner, G. E. P, (1996). Multivariate logistic models for incomplete binary responses. *J. Am. Statist. Assoc*. **91**, 99–108.

Flehinger, B. J., Reiser, B., and Yashchin, E. (1996). Inference about defects in the presence of masking. *Technometrics* **38**, 247–255.

Flury, B. D. (1988). *Common Principal Components and Relater Multivariate Models*. Wiley, New York.*

Flury, B. D. (1997). *A First Course in Multivariate Statistics*. Springer-Verlag, New York.*

Flury, B. D., Nel, D. G., and Pienaar, I. (1995). Simultaneous detection of shift in means and variances. *J. Am. Statist. Assoc*. **90**, 1474–1481.

Fong, D. K. H., and Bolton, G. E. (1997). Analyzing ultimatum bargaining: A Bayesian approach to the comparison of two potency curves under shape constraints. *J. Bus. Econ. Statist*. **15**, 335–344.

Forster, J. J., McDonald, J. W., and Smith, P. W. F. (1996). Monte Carlo exact conditional tests for log-linear and logistic models. *J. R. Statist. Soc. B* **58**, 445–453.

Fortin, V., Bernier, J., and Bobee, B. (1997). Simuation, Bayes, and bootstrap in statistical hydrology. *Water Resour. Res*. **33**, 439–448.

Foster, D. H., and Bischof, W. F. (1997). Bootstrap estimates of the statistical accuracy of thresholds obtained from psychometric functions. *Spatial Vision* **11**, 135–139.

Foutz, R. V. (1980). A method for constructing exact tests from test statistics that have unknown null distributions. *J. Statist. Comput. Simul*. **10**, 187–193.

Frangos, C. C., and Schucany, W. R. (1990). Jackknife estimation of the bootstrap acceleration constant. *Comp. Statist. Data Anal*. **9**, 271–282.*

Frangos, C. C., and Schucany, W. R. (1995). Improved bootstrap confidence intervals in certain toxicological experiments. *Commun. Statist. Theory Methods* **24**, 829–844.

Frangos, C. C., and Stone, M. (1984). On jackknife cross-validatory and classical methods of estimating a proportion with batches of different sizes. *Biometrika* **71**, 361–366.

Frangos, C. C., and Swanepoel, C. J. (1994). Bootstrap confidence intervals for a slope parameter of a logistic model. *Commun. Statist. Simul. Comput*. **23**, 1115–1126.

Franke, J., and Hardle, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist*. **20**, 121–145.

Franke, J., and Wendel, M. (1992). A bootstrap approach to nonlinear autoregression: Some preliminary results. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG* (K. H. Jockel, G. Rothe, and W. Sendler, editors), *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 101–105. Springer-Verlag, Berlin.

Franklin, L. A., and Wasserman, G. S. (1991). Bootstrap confidence interval estimates of C_{pk}: An introduction. *Commun. Statist. Simul. Comput*. **20**, 231–242.*

Franklin, L. A., and Wasserman, G. S. (1992). Bootstrap lower confidence limits for capability indices. *J. Qual. Technol*. **24**, 196–210.*

Franklin, L. A., and Wasserman, G. S. (1994). Bootstrap lower confidence limit estimates of C_{jkp} (the new flexible process capability index). *Pakistan J. Statist. A* **10**, 33–45.*

Freedman, D. A. (1981), Bootstrapping regression models. *Ann. Statist*. **9**, 1218–1228.*

Freedman, D. A. (1984). On bootstrapping two-stage least squares estimates in stationary linear models. *Ann. Statist*. **12**, 827–842.*

Freedman, D. A. Navidi, W., and Peters, S. C. (1988). On the impact of variable selection in fitting regression equations. In *Old Model Uncertainty and Its Statistical Implications* (T. K. Dijkstra, editor), *Lecture Notes in Economics and Mathematical Systems*, Vol. 306. Springer-Verlag, Berlin.

Freedman, D, A., and Peters, S. C. (1984a). Bootstrapping a regression equation: Some empirical results. *J. Am. Statist. Assoc*. **79**, 97–106,*

Freedman, D, A., and Peters, S. C. (1984b). Bootstrapping an econometric model: Some empirical results. *J. Bus. Econ. Statist*. **2**, 150–158.*

Fresen, J. L., and Fresen, J. W. (1986). Estimating the parameter in the Pauling equation. *J. Appl. Statist*. **13**, 27–37.

Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist*. **19**, 1–141.

Friedman, J. H., and Stuetzle, W. (1981). Projection pursuit regression. *J. Ann. Statist. Assoc*. **76**, 817–823.

Friedman, L. W., and Friedman, H. H. (1995). Analyzing simulation output using the bootstrap method. *Simulation* **64**, 95–100.

Fuchs, C. (1978). On test sizes in linear models for transformed variables. *Technometrics* **20**, 291–299.*

Fujikoshi, Y. (1994). On the bootstrap approximations for Hotelling's T^2 statistic. In *5th Japan–China Symposium on Statistics* (M. Ichimura, S. Mao, and G, Fan, editors), Vol. 5, pp. 69–71.

Fukuchi, J. I. (1994). Bootstrapping extremes of random variables. Ph.D. Dissertation, Iowa State University, Ames.*

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, 2nd ed. Academic Press, San Diego.*

Fukunaga, K., and Hayes, R. R. (1989). Estimation of classifier performance. *IEEE Trans. Pattern Anal. Mach. Intell*. **PAMI-11**, 1087–1101.

Fuller, W. A. (1976). *Introduction to Statistical Tim Series*. Wiley, New York.

Furlanello, C., Merer, S., Chemini, C., and Rizzoli, A. (1998). An applicationof the bootstrap 632+ rule to ecological data. In *Proceedings of the 9th Italian Workshop on Neural Nets* (M. Marinaro, and R. Talgliaferri, editors), pp. 227–232.*

Gabriel, K. R., and Hsu, C. F. (1983). Evaluation of the power of rerandomization tests, with application to weather modification experiments. *J. Am. Statist. Assoc*. **78**, 766–775.

Gaenssler, P. (1987). Bootstrapping empirical measures indexed by Vapnik–Chervonenkis classes of sets. In *Probability Theory and Mathematical Statistics*. pp. 467–481, VNU Science Press, Utrecht.

Gaenssler, P. (1992). Confidence bands for probability distributions on Vapnik–Chervonenkis classes of sets in arbitrary sample spaces using the bootstrap. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG* (K.-H. Jockel, G,

Rothe, and W. Sendler, editors), *Lecture Notes in Economics and Mathematical Systems*, 376, pp. 57–61. Springer-Verlag, Berlin.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York.*

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed. Krieger, Malabar.*

Gallant, A. R. (1987). *Nonlinear Statistical Models*. Wiley, New York.*

Ganeshanandam, S., and Krzanowski, W. J. (1989). On selecting Variables and assessing their performance inlinear discriminant analysis. *Aust. J. Statist*. **31**, 433–448.

Ganeshanandam, S., and Krzanowski, W. J. (1990). Error-rate estimation in two-group discriminant analysis using linear discriminant functions. *J. Statist. Comput. Simul*. **36**, 157–176.

Gangopadhyay, A. K., and Sen, P. K. (1990). Bootstrap confidence intervals for conditional quantile functions. *Sankhya A* **52**, 346–363.

Ganoe, F. J. (1989). Statistical bootstrap with computer simulation: Methodology for fuzzy-logic-based expert system validation. In *Proceedings of the 18th Annual Western Regional Meeting of the Decision Sciences Institute* (V. V. Bellur and J. C. Rogers, editors), Vol. 18, pp. 211–213. Decision Sciences Institute.

Garcia-Cortes, L. A., Moreno, C., Varona, L., and Altarriba, J. (1995). Estimation of prediction-error variances by resampling. *J. Anim. Breeding Genet*. **112**, 176–182.

Garcia-Jurado, I., Gonzalez-Manteiga, W., Prada-Sanchez, J. M., Febrero-Bande, M., and Cao. R. (1995). Predicting using Box-Jenkins, nonparametric, and bootstrap techniques. *Technometrics* **37**, 303–310.

Garcia-Soidan, P. H., and Hall, P. (1997). On sample reuse methods for spatial data. *Biometrics* **53**, 273–281.

Garthwaite, P. H., and Buckland, S. T. (1992). Generating Monte Carlo confidence intervals by the Robbins–Monro process. *Appl. Statist*. **41**, 159–171.

Gatto, R., and Ronchetti, E. (1996). General saddlepoint approximations of marginal densities and tail probabilities. *J. Am. Statist. Assoc*. **91**, 666–673.

Gaver, D. P., and Jacobs, P. A. (1989). System availability: Time dependence and statistical inference by (semi) non-parametric methods. *Appl. Stach. Models Data Anal*. **5**, 357–375.

Geisser, S. (1975). The predictive sample reuse method with applications. *J. Am. Statist. Assoc*. **70**, 320–328.

Geisser, S. (1993). *Predictive Inference: An Introduction*. Chapman & Hall, London.*

Geissler, P. H. (1987). Bootstrapping the lognormal distribution. *Proc. Comput. Sci. Statist*. **19**, 543–545.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.*

Gentle, J. E. (1985). Monte Carlo methods. In *Encyclopedia of Statistical Sciences*, Vol. 5, pp. 612–617, Wiley, New York.

George, P., J., Oksanen, E. H., and Veall, M. R. (1995). Analytic and bootstrap approaches to testing a market saturation hypothesis. *Math. Comput. Simul*. **39**, 311–315.

George, S. L. (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Statist. Med*. **4**, 39–46.

Geweke, J. (1993). Inference and forecasting for chaotic non-linear time series. In *Nonlinear Dynamics and Evolutionary Economics* (P. Chen, and R. Day, editors). Oxford University Press, Oxford.

Geyer, C. J. (1991). Markov Chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: Proceedings of 23rd Symposium on the Interface* (E. M. Keramidas, editor), pp. 156–163. Interface Foundation, Fairfax.

Geyer, C. J. (1995a). Likelihood ratio tests and inequality constraints. Technical Report 610, University of Minnesota, School of Statistics.

Geyer, C. J. (1995b). Estimation and optimization of functions. In *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors), pp. 241–258. Chapman & Hall, London.

Geyer, C. J., and Moller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.* **21**, 359–373.

Ghorbel, F., and Banga, C. (1994). Bootstrap sampling applied to image analysis. In Proceedings of the 1994 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 6, pp. 81–84.

Ghosh, M. (1985). Berry-Essen bounds for functions of *U*-statistics. *Sankhya A* **47**, 255–270.

Ghosh, M., and Meeden, G. (1997). *Bayesian Methods for Finite Population Sampling*. Chapman & Hall, London.*

Ghosh, M., Parr, W. C., Singh, K., and Babu, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12**, 1130–1135.

Giachin, E. Baggia, P., and Micca, G. (1994). Language models for spontaneous speech recognition: A bootstrap method for learning phrase bigrams. In *Proceedings of the 1994 Conference on Spoken Language Processing*. Vol. 2, pp. 843–846.

Gifi, A. (name of a group of Dutch statisticians) (1990) *Nonlinear Multivariate Analysis*. Wiley, Chichester.*

Gigli, A. (1994a). Contributions to importance sampling and resampling. Ph.D. Thesis, Department of Mathematics, Imperial College, London.*

Gigli, A. (1994b). Efficient bootstrap methods: A review. Consiglio Nazionale delle Ricerche, Rome. Technical Report Number Quarderno-7.*

Gine, E. (1997). Lectures on some aspects of the bootstrap theory. In *Summer School of Probability* (E. Gine, G. R. Grimmett, and L. Saloff-Coste, editors), *Lecture Notes on Mathematics*, Vol. 1665, pp. 37–152. Springer-Verlag, New York.

Gine, E., and Zinn, J. (1989). Necessary conditions for bootstrap of the mean. *Ann. Statist.* **17**, 684–691.*

Gine, E., and Zinn, J. (1990). Bootstrap general empirical measures. *Ann. Probab.* **18**, 851–869.

Gine, E., and Zinn, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Ann. Probab.* **19**, 758–782.

Glasbey, C. A. (1987). Tolerance-distribution-free analysis of quantal dose–response data. *Appl. Statist.* **236**, 252–259.

Gleason, J. R. (1988). Algorithms for balanced bootstrap simulations. *Am. Statist.* **42**, 263–266.

Gleser, L. J. (1985). A note on G. R. Dolby's unreplicated ultrastructural model. *Biometrika* **72**, 117–124.

Glick, N. (1978). Additive estimators for probabilities of correct classification. *Pattern Recogn*. **10**, 211–222.*

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysus of Multivariate Observations*. Wiley, New York.*

Gnanadesikan, R. (1997). *Methods for Statistical Data Analysus of Multivariate Observations*, 2nd ed., Wiley, New York.*

Gnanadesikan, R., and Kettenring, J. R. (1982). Data-based metrics for cluster analysis. *Util. Mat.* **21A**, 75–99.

Golbeck, A. L. (1992). Bootstrapping current life table estimators. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG* (K-H. Jockel, G. Roth, and W. Sendler, editors), *Lecture Notes in Economics and Mathematical Systems*, Vol. 376, pp. 197–201. Springer-Verlag, Berlin.

Goldstein, H. (1995). *Kendall's Library of Statistics 3: Multilevel Statistical Models*, 2nd ed., Edward Arnold, London.*

Gong, G. (1982). Some Ideas to using the bootstrap in assessing model variability in regression. *Proc. Comput. Sci. Statist*. **24**, 169–173.*

Gong, G. (1986). Cross-validation, the jackknife, and the bootstrap: Excess error in forward logistic regression. *J. Am. Statist. Assoc*. **81**, 108–113.*

Gonzalez, L., and Manly, B. F. J. (1993). Bootstrapping for sample design with quality surveys. ASA Proceedings of Quality and Productivity, pp. 262–265.

Gonzalez-Manteiga, W., Prada-Sanchez, J. M., and Romo, J. (1993). The bootstrap—a review. *Comput. Statist. Q*. **9**, 165–205.*

Good, P. (1989). Almost most powerful tests for composite alternatives. *Commun. Statist. Theory Methods* **18**, 1913–1925.

Good, P. (1994). *Permutation Tests*. Springer-Verlag, New York.*

Good, P. (1998). *Resampling Methods: A Practical Guide to Data Analysis*. Birkhauser, Boston.*

Good, P., and Chernick, M. R. (1993). Testing the equality of variances of two populations. Unpublished manuscript.*

Goodnight, C. J., and Schwartz, J. M. (1997). A bootstrap comparison of genetic covariance matrices. *Biometrics* **53**, 1026–1039.

Götze, F., and Künsch, H. R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist*. **24**, 1914–1933.*

Gould, W. R., and Pollock, K. H. (1997). Catch-effort estimation of population parameters under the robust design. *Biometrics* **53**, 207–216.

Graham, R. L., Hinkley, D. V., John, P. W. M., and Shi, S. (1990). Balanced design of bootstrap simulations. *J. R. Statist. Soc. B* **52**, 185–202.*

Graubard, B. I., and Korn. E. L. (1993). Hypothesis testing with complex survey data: The use of classical quadratic test statistics with particular reference to regression problems. *J. Am. Statist. Assoc*. **88**, 629–641.

Gray, H. L., and Schucany, W. R. (1972). *The Generalized Jackknife Statistic*. Marcel Dekker, New York.*

Green, P. J., and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman & Hall, London.

Green, R., Hahn, W., and Rocke, D. (1987). Standard errors for elasticities: A comparison of bootstrap and asymptotic standard errors. *J. Bus. Econ. Statist*. **5**, 145–149.*

Greenacre. M. J. (1984). *Theory and Application of Correspondence Analysis*. Academic Press, London.*

Gross, S. (1980). Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods*. American Statistical Association, Alexandria.*

Gross, S. T., and Lai, T. L. (1996a). Bootstrap methods for truncated and censored data. *Statist. Sin*. **6**, 509–530.*

Gross, S. T., and Lai, T. L. (1996b). Nonparametric estimators and regression analysis wioth left-truncated and right-censored data. *J. Am. Statist. Assoc*. **91**, 1166–1180.

Gruet, M. A., Huet, S., and Jolivet, E. (1993). Practical use of bootstrap in regression. In *Computer Intensive Methods in Statistics*, pp. 150–166, *Statist. Comp. Physica*, Heidelberg.

Gu, C. (1987). What happens when bootstrapping the smoothing spline? *Commun. Statist. Theory Methods* **16**, 3275–3284.

Gu, C. (1992). On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. *Can. J. Statist*. **20**, 399–414.

Guan, Z. (1993). The bootstrap of estimators of symmetric distribution functions. *Chin. J. Appl. Probab. Statist*. **9**, 402–408.

Guerra, R., Polansky, A. M., and Schucany, W. R. (1997). Smoothed bootstrap confidence intervals with discrete data. *Comput. Statist. Data Anal*. **26**, 163–176.

Guillou, A. (1995). Weighted bootstraps for studentized statistics. *C. R. Acad. Sci. Paris Ser. I Math* **320**, 1379–1384.

Gunter, B. H. (1989a). The use and abuse of C_{pk}, *Qual. Prog*. **22** (**3**), 108–109.*

Gunter, B. H. (1989b). The use and abuse of C_{pk}, *Qual. Prog*. **22** (**5**), 79–80.*

Haeusler, E., Mason, D. M., and Newton, M. A. (1992). Weighted bootstrapping of means. *Cent. Wisk. Inf. Q*. **5**, 213–228.

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*. Wiley, New York.*

Hall, P. (1983). Inverting an Edgeworth expansion. *Ann. Statist*. **11**, 569–576.

Hall, P. (1985). Resampling a coverage pattern. *Stoc. Proc*. **20**, 231–246.*

Hall, P. (1986a). On the bootstrap and confidence intervals. *Ann. Statist*. **14**, 1431–1452.*

Hall, P. (1986b). On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist*. **14**, 1453–1462.*

Hall, P. (1987a). On bootstrap and likelihood-based confidence regions. *Biometrika* **74**, 481–493.

Hall, P. (1987b). On the bootstrap and continuity correction. *J. Roy. Statist. Soc. B* **49**, 82–89.

Hall, P. (1987c). Edgeworth expansion for Student's *t* statistic under minimal moment conditions. *Ann. Probab*. **15**, 920–931.

Hall, P. (1988a). On the bootstrap and symmetric confidence intervals. *J. Roy. Statist. Soc. B* **50**, 35–45.

Hall, P. (1988b). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist*. **16**, 927–985.*

Hall, P. (1988c). *Introduction to the Theory of Coverage Processes*. Wiley, New York.*

Hall, P. (1988d). Rate of convergence in bootstrap approximations. *Ann. Probab*. **16**, 1665–1684.*

Hall, P. (1989a). On efficient bootstrap simulation. *Biometrika* **76**, 613–617.*

Hall, P. (1989b). Antithetic resampling for the bootstrap. *Biometrika* **76**, 713–724.*

Hall, P. (1989c). Unusual properties of bootstrap confidence intervals in regression problems. *Probab. Th. Rel. Fields* **81**, 247–274.*

Hall, P. (1989d). On convergence rates in nonparametric problems. *Int. Statist. Rev*. **57**, 45–58.

Hall, P. (1990a). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist*. **18**, 121–140.

Hall, P. (1990b). Using the bootstrap to estimate mean squared error and select smoothing parameters in nonparametric problems. *J. Mult. Anal*. **32**, 177–203.

Hall, P. (1990c). Performance of bootstrap balanced resampling in distribution function and quantile problems. *Probab. Theory Relat. Fields* **85**, 239–260.

Hall, P. (1990d). Asymptotic properties of the bootstrap for heavy-tailed distributions. *Ann. Probab*. **18**, 1342–1360.

Hall, P. (1991a). Bahadur representations for uniform resampling and importance resampling, with applications to asymptotic relative efficiency. *Ann. Statist*. **19**, 1062–1072.*

Hall, P. (1991b). On relative performance of bootstrap and Edgeworth approximations of a distribution function. *J. Multivar. Anal*. **35**, 108–129.

Hall, P. (1991c). Balanced importance resampling for the bootstrap. Unpublished manuscript.

Hall, P. (1991d). On bootstrap confidence intervals in nonparametric regression. Unpublished manuscript.

Hall, P. (1991e). On Edgeworth expansions and bootstrap confidence bands in nonparametric curve estimation. Unpublished manuscript.

Hall, P. (1991f). Edgeworth expansions for nonparametric density estimators, with applications to asymptotic relative efficiency. *Ann. Statist*. **19**, 1062–1072.

Hall, P. (1991g). Edgeworth expansions for nonparametric density estimators with applications. *Statistics* **22**, 215–232.

Hall, P. (1992a). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.*

Hall, P. (1992b). On the removal of skewness by transformation. *J. Roy. Statist. Soc. B* **54**, 221–228.

Hall, P. (1992c). Efficient bootstrap simulations. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard editors), pp. 127–143. Wiley, New York.*

Hall, P. (1992d). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist*. **20**, 695–711.

Hall, P. (1992e). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist*. **20**, 675–694.

Hall, P. (1994). A short history of the bootstrap. In *Proceedings of the 1994 International Conference on Acoustics, Speech and Signal Processing*. **6**, 65–68.*

Hall, P. (1995). On the biases of error estimators in prediction problems. *Statist. Probab. Lett*. **24**, 257–262.

Hall, P. (1997). Defining and measuring long-range dependence. In *Nonlinear Dynamics and Time Series*. (C. D. Cutler and D. T. Kaplan, editors), Vol. 11, *Fields Communications*, pp. 153–160.*

Hall, P. (1998). Block bootstrap. In *Encyclopedia of Statistical Sciences*, Update Volume 2 (S. Kotz, C. B. Read, and D. L. Banks, editors), pp. 83–84. Wiley, New York.*

Hall, P., DiCiccio, T. J., and Romano, J. P. (1989). On smoothing and the bootstrap. *Ann. Statist*. **17**, 692–704.

Hall, P., Hardle, W., and Simar, L. (1993). On the inconsistency of bootstrap distribution estimators. *Comput. Statist. Data Anal*. **16**, 11–18.*

Hall, P., and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Am. Statist. Assoc*. **85**, 1039–1049.

Hall, P., and Horowitz, J. L. (1993). Corrections and blocking rules for the block bootstrap with dependent data. Technical Report SR11–93, Centre for Mathematics and Its Applications, Australian National University.

Hall, P., and Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized method of moments estimators. *Econ*. **64**, 891–916.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561–574.*

Hall, P., Huber, C., and Speckman, P. L. (1997). Covariate-matched one-sided tests for the difference between functional means. *J. Am. Statist. Assoc*. **92**, 1074–1083.

Hall, P., and Jing, B. Y. (1996). On sample reuse methods for dependent data. *J. Roy. Statist. Soc.B* **58**, 727–737.*

Hall, P., Jing, B. Y., and Lahiri, S. N. (1998). *Statist. Sin*. **8**, 1189–1204.

Hall, P., and Keenan, D. M. (1989). Bootstrap methods for constructing confidence regions for hands. *Communs. Statist. Stochastic Models* **5**, 555–562.

Hall, P., and LaScala, B. (1990). Methodology and algorithms of empirical likelihood. *Int. Statist. Rev*. **58**, 109–127.

Hall, P., and Martin, M. A. (1988a). On bootstrap resampling and iteration. *Biometrika* **75**, 661–671.*

Hall, P., and Martin, M. A. (1988b). On the bootstrap and two-sample problems. *Austral. J. Statist*. **30A**, 179–192.

Hall, P., and Martin, M. A. (1989a). Exact convergence rate of the bootstrap quantile variance estimator. *Probab. Th. Rel. Fields* **80**, 261–268.

Hall, P., and Martin, M. A. (1989b). A note on the accuracy of bootstrap percentile method confidence intervals for a quantile. *Statist. Probab. Lett*. **8**, 197–200.

Hall, P., and Martin, M. A. (1991). On the error incurred using the bootstrap variance estimate when constructing confidence intervals for quantiles. *J. Multivar. Anal*. **38**, 70–81.

Hall, P., Martin, M. A., and Schucany, W. R. (1989). Better nonparametric bootstrap confidence intervals for the correlation coefficient. *J. Statist. Comput. Simul*. **33**, 161–172.*

Hall, P., and Mammen, E. (1994). On general resampling algorithms and their performance in distribution estimation. *Ann. Statist*. **22**, 2011–2030.

Hall, P., and Owen, A. B. (1989). Empirical likelihood confidence bands in curve estimation. Unpublished manuscript.

Hall, P., and Owen, A. B. (1993). Empirical likelihood confidence bands in density estimation. *J. Comput. Graphical Statist*. **2**, 273–289.

Hall, P., and Padmanabhan, A. R. (1997). Adaptive inference for the two-sample scale problem. *Technometrics* **39**, 412–422.*

Hall, P., and Pittelkow, Y. E. (1990). Simultaneous bootstrap confidence bands in regression. *J. Statist. Comput. Simul*. **37**, 99–113.

Hall, P., and Sheather, S. J. (1988). On the distribution of a studentized quantile. *J. Roy. Statist. Soc. B* **50**, 381–391.

Hall, P., and Titterington, D. M. (1989). The effect of simulation order on level accuracy and power of Monte Carlo tests. *J. R. Statist. Soc. B* **51**, 459–467.

Hall, P., and Weissman, I. (1997). On the estimation of extreme tail probabilities. *Ann. Statist*. **25**, 1311–1326.

Hall, P., and Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics* **47**, 757–762.*

Hall, P., and Wolff, R. C. L. (1995). Properties of invariant distributions and Lyapunov exponents for chaotic logistic maps. *J. R. Statist. Soc. B* **57**, 439–452.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.*

Hamilton, M. A., and Collings, B. J. (1991). Determining the appropriate sample size for nonparametric tests for location shift. *Technometrics* **33**, 327–337.

Hammersley, J. M., and Handscomb, D. C. (1964). *Monte Carlo Methods*. Methuen, London.*

Hammersley, J. M., and Morton, K. W. (1956). A new Monte Carlo technique: Antithetic variates. *Proc. Cambridge Philos. Soc*. **52**, 449–475.*

Hampel, F. R. (1973). Some small-sample asymptotics. In *Proceedings of the Prague Symposium on Asymptotic Statistics* (J. Hajek, editor), Charles University, Prague.*

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Am. Statist. Assoc*. **69**, 383–393.*

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.*

Hand, D. J. (1981). *Discrimination and Classification*. Wiley, Chichester.*

Hand, D. J. (1982). *Kernel Discriminant Analysis*. Wiley, Chichester.*

Hand, D. J. (1986). Recent advances in error rate estimation. Pattern Recogn. *Lett*. **4**, 335–340.*

Hardle, W. (1989). Resampling for inference from curves. *Proc. 47th Session ISI*, **4**, 53–54, Paris.

Hardle, W. (1990a). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.*

Hardle, W. (1990b). *Smoothing Techniques with Implementation in S*. Springer-Verlag, New York.*

Hardle, W., and Bowman, A.W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Am. Statist. Assoc*. **83**, 102–110.*

Hardle, W., Hall, P. and Marron, S. (1988). How far are automatically chosen smoothing parameters from their optimum? (with discussion). *J. Am. Statist. Assoc*. **83**, 86–101.

Hardle, W., Huet, S., and Jolivet, E. (1990). Better bootstrap confidence intervals for regression curve estimation. Unpublished manuscript.

Hardle, W., and Kelly, G. (1987). Nonparametric kernel regression estimation—Optimal choice of bandwidth. *Statistics* **18**, 21–35.

Hardle, W., and Mammen, E. (1991). Bootstrap methods in nonparametric regression. In *Nonparametric Functional Estimation and Related Topics. Proceedings of the NATO Advanced Study Institute*, Spetses, Greece (G. Roussas, editor).

Hardle, W., and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist*. **13**, 1465–1481.

Hardle, W., and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist*. **19**, 778–796.*

Hardle, W., Marron, J. S., and Wand, M. P. (1990). Bandwidth choice for densiity derivatives. *J. R. Statist. Soc. B* **52**, 223–232.

Hardle, W., and Nussbaum, M. (1992). Bootstrap Confidence Bands. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe and W. Sendler editors), Vol. 376, pp. 63–70. Springer-Verlag, Berlin.

Harrell, F. E., and Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika* **69**, 635–640.

Harshman, J. (1994). The effects of irrelevant characters on bootstrap values. *Syst. Biol*. **43**, 419–424.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer-Verlag, New York.*

Hartigan, J. A. (1969). Using subsample values as typical values. *J. Am. Statist. Assoc*. **64**, 1303–1317.*

Hartigan, J. A. (1971). Error analysis by replaced samples. *J. R. Statist. Soc. B* **33**, 98–1130.*

Hartigan, J. A. (1975). Necessary and sufficient conditions for the asymptotic joint normality of a statistic and its subsample values. *Ann. Statist*. **3**, 573–580.*

Hartigan, J. A. (1990). Perturbed periodogram estimates of variance. *Int. Statist. Rev*. **58**, 1–7.

Hartigan, J. A., and Forsythe, A. (1970). Efficiency and confidence intervals generated by repeated subsample calculations. *Biometrika* **57**, 629–640.

Hasegawa, M., and Kishino, H. (1994). Accuracies of the simple methods for estimating the bootstrap probability of a maximum-likelihood tree. *Mol. Biol. Evol*. **11**, 142–145.

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics* **8**, 431–444.*

Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *J. Am. Statist. Assoc*. **64**, 1459–1471.*

Hastie. T. J., and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall, London.*

Hauck, W.W., and Anderson, S. (1991). Individual bioequivalence: what matters to the patient. *Statist. Med*. **10**, 959–960.*

Hauck, W. W., McKee, L. J., and Turner, B. J. (1997). Two-part survival models applied to administrative data for determining rate of and predictors for maternal–child transmission of HIV. *Statist. Med*. **16**, 1683–1694.

Haukka, J. K. (1995). Correction for covariate measurement error in generalized linear models—a bootstrap approach. *Biometrics* **51**, 1127–1132.

Hawkins, D. M. editor (1982). *Topics in Applied Multivariate Analysis*. Cambridge University Press, Cambridge.*

Hawkins, D. M., Simonoff, J. S., and Stromberg, A. J. (1994). Distributing a computationally intensive estimator: The case of exact LMS regression. *Comput. Statist*. **9**, 83–95.

Hayes, K. G., Perl, M. L., and Efron, B. (1989). Applications of the bootstrap statistical method to the tau-decay-mode problem. *Phys. Rev. D* **39**, 274–279.*

He, K. (1987). Bootstrapping linear M-regression models. *Acta Math. Sin*. **29**, 613–617.

Heavlin, W. D. (1988). Statistical properties of capability indices. Technical Report # 320, Advanced Micro Devices, Inc., Sunnyvale, CA.

Heimann, G., and Kreiss, J.-P. (1996). Bootstrapping general first order autoregression. *Statist. Probab. Lett*. **30**, 87–98.*

Heitjan, D. F., and Landis, J. R. (1994). Assessing secular trends in blood pressure: A multiple-imputation approach. *J. Am. Statist. Assoc*. **89**, 750–759.

Heller, G., and Venkatraman, E. S. (1996). Resampling procedures to compare two survival distributions in the presence of right-censored data. *Biometrics* **52**, 1204–1213.

Helmers, R. (1991a). On the Edgeworth expansion and the bootstrap approximation for a studentized *U*-statistic. *Ann. Statist*. **19**, 470–484.

Helmers, R. (1991b). Bootstrap Methods. Unpublished manuscript.*

Helmers, R., Janssen, P., and Serfling, R. (1988). Glivenko–Cantelli properties of some generalized empirical DF's and strong convergence of generalized *L*-statistics. *Probab. Th. Rel. Fields* **79**, 75–93.

Helmers, R., Janssen, P., and Serfling, R. (1990). Berry-Esseen and bootstrap results for generalized *L*-statistics. *Scand. J. Statist*. **17**, 65–78.

Helmers, R., Janssen, P., and Veraverbeke, N. (1992). Bootstrapping *U*-quantiles. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 145–155. Wiley, New York.*

Hesterberg, T. (1988). Variance reduction techniques for bootstrap and other Monte Carlo simulations. Ph.D. dissertation, Department of Statistics, Stanford University.*

Hesterberg, T. (1992). Efficient bootstrap simulations I: Importance sampling and control variates. Unpublished manuscript.

Hesterberg, T. (1995a). Tail-specific linear approximations for efficient bootstrap simulations. *J. Comput. Graph. Statist*. **4**, 113–133.*

Hesterberg, T. (1995b). Weighted average importance sampling and defensive mixture distributions. *Technometrics* **37**, 185–194.*

Hesterberg, T. (1996). Control variates and importance sampling for efficient bootstrap simulations. *Statist. Comput*. **6**, 147–157.*

Hesterberg, T. (1997). Fast bootstrapping by combining importance sampling and concomitants. *Proc. Comput. Sci. Statist*. **29**, 72–78.*

Hewer, G., Kuo, W., and Peterson, L. (1996a). Multiresolution detection of small objects using bootstrap methods and wavelets. In *Proceedings of the Conference on Signal and Data Processing of Small Targets, Orlando, FL, April 9–11, 1996. SPIE Proceedings*, Vol. 2759, pp. 2–10.

Hewer, G., Kuo, W., and Peterson, L. (1996b). Adaptive wavelet detection of transients using the bootstrap. *Proc. SPIE* **2762**, 105–114.

Higgins, K. M., Davidian, M., and Giltinan, D. M. (1997). A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to IGF-I pharmacokinetics. *J. Am. Statist. Assoc*. **92**, 436–448.

Hill, J. R. (1990). A general framework for model-based statistics. *Biometrika* **77**, 115–126.

Hillis, D. M., and Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol*. **42**, 182–192.

Hills, M. (1966). Allocation rules and their error rates. *J. R. Statist. Soc. B* **28**, 1–31.*

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics* **19**, 285–292.

Hinkley, D. V. (1983). Jackknife Methods. In *Encyclopedia of Statistical Sciences*, Vol. 4, pp. 280–287. Wiley, New York.

Hinkley, D. V. (1984). A hitchhiker's guide to the galaxy of theoretical statistics. In *Statistics: An Appraisal* (H. A. David and H. T. David, editors), pp. 437–453. Iowa State University Press, Ames.*

Hinkley, D. V. (1988). Bootstrap methods (with discussion). *J. R. Statist. Soc. B* **50**, 321–337.*

Hinkley, D. V. (1989). Bootstrap significance tests. *Proceedings of the 47th Session of International Statistics Institute, Paris*, pp. 65–74.

Hinkley, D. V., and Schechtman, E. (1987). Conditional bootstrap methods in the mean-shift model. *Biometrika* **74**, 85–94.

Hinkley, D. V., and Shi, S. (1989). Importance sampling and the bootstrap. *Biometrika* **76**, 435–446.*

Hinkley, D. V., and Wei, B. C. (1984). Improvements of jackknife confidence limit methods. *Biometrika* **71**, 331–339.

Hirst, D. (1996). Error-rate estimation in multiple-group linear discriminant analysis. *Technometrics* **38**, 389–399.*

Hjort, N. L. (1985). Bootstrapping Cox's regression model. Technical Report NSF-241, Department of Statistics, Stanford University.

Hjort, N. L. (1992). On inference in parametric data models. *Int. Statist. Rev*. **60**, 355–387.

Hjorth, J. S. U. (1994). *Computer Intensive Statistical Methods: Validation, Model Selection and Bootstrap*. Chapman & Hall, London.*

Holbert, D., and Son, M.-S. (1986). Bootstrapping a time series model: Some empirical results. *Commun. Statist. Theory Methods* **15**, 3669–3691.*

Holm, S. (1993). Abstract bootstrap confidence intervals in linear models. *Scand. J. Statist*. **20**, 157–170.

Hoffman, W. P., and Leurgans, S. E. (1990). Large sample properties of two tests for independent joint action of two drugs. *Ann. Statist*. **18**, 1634–1650.

Hope, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc. B*. 582–598.*

Horowitz, J. L. (1994). Bootstrap-based critical values for the information matrix test. *J. Econometrics* **61**, 395–411.

Horvath, L., and Yandell, B. S. (1987). Convergence rates for the bootstrapped product limit process. *Ann. Statist*. **15**, 1155–1173.

Hsieh, D. A., and Manski, C. F. (1987). Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Ann. Statist*. **15**, 541–551.

Hsieh, J. J. (1992). A hazard process for survival analysis. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 345–362. Wiley, New York.*

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. Chapman & Hall, New York.*

Hsu, Y.-S., Walker, J. J., and Ogren, D. E. (1986). A stepwise method for determining the number of components in a mixture. *Math. Geog*. **18**, 153–160.

Hu, F. (1997). Estimating equations and the bootstrap. In *Estimating Functions* (I. V. Basawa, V. P. Godambe, and R. L. Taylor, editors), *Institute of Mathematical Statistics Monograph Series*, Vol. 32, pp. 405–416.

Hu, F., and Zidek, J. V. (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika* **82**, 263–275.

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**, 45–66.

Huang, J. S., Sen, P. K., and Shao, J. (1995). Bootstrapping a sample quantile when the density has a jump. *Statist. Sin*. **6**, 299–309.

Huang, X., Chen, S., and Soong, S.-J. (1998). Piecewise exponential survival trees with time-dependent covariates. *Biometrics* **54**, 1420–1433.

Hubbard, A. E., and Gilinsky, N. L. (1992). Mass extinctions as statistical phenomena: An examination of the evidence using χ^2 tests and bootstrapping. *Paleobiology* **18**, 148–160.

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.*

Huet, S., and Jolivet, E. (1989). Exactitude au second order des intervalles de cofiance bootstrap pour les parametres d'un modele de regression non lineare. *C. R. Acad. Sci. Paris Ser. I. Math.* **308**, 429–432.

Huet, S., Jolivet, E., and Messean, A. (1990). Some simulation results about confidence intervals and bootstrap methods in nonlinear regression. *Statistics* **21**, 369–432.

Hur, K., Oprian, C. A., Henderson, W. G., Thakkar, B., and Urbanski, S. (1996). A SAS/sup (R/) macro for validating a logistics model with a split sample and bootstrap methods. *SUGI 21*, **1**, 953–956.

Hurvich, C. M., Simonoff, J. S., and Zeger, S. L. (1991). Variance estimation for sample autocovariances. Direct and resampling approaches. *Aust. J. Statist.* **33**, 23–42.

Hurvich, C. M., and Tsai, C.-L. (1990). The impact of model selection on inference in linear regression. *Am. Statist.* **44**, 214–217.

Hurvich, C. M., and Zeger, S. L. (1987). Frequency domain bootstrap methods for time series *Statistics and Operations Research* working paper. New York University, New York.*

Huskova, M., and Janssen, P. (1993a). Generalized bootstrap for studentized *U*-statistics: a rank statistic approach. *Statist. Probab. Lett.* **16**, 225–233.

Huskova, M., and Janssen, P. (1993b). Consistency of the generalized bootstrap for degenerate *U*-statistics. *Ann. Statist.* **21**, 1811–1823.

Hwa-Tung, O., and Zoubir, A. M. (1997). Non-Gaussian signal detection from multiple sensors using the bootstrap. In *Proceedings of ICICS, 1997 International Conference on Information, Communications, and Signal Processing: Trends in Information Systems Engineering and Wireless Multimedia Communications*, Vol. 1, pp. 340–344.

Hyde, J. (1980) Testing survival with incomplete observations. In *Biostatistics Casebook* (R. G. Miller Jr., B. Etron, B. Brown, and L. Moses, editors), Wiley, New York.*

Iglewicz, B., and Shen, C. F. (1994). Robust and bootstrap testing procedures for bioequivalence. *J. Biopharm. Statist.* **4**, 65–90.

Izenman, A. J. (1985). Bootstrapping Kolmogorov–Smirnov statistics. *Proc. ASA Sect. Statist. Comp.*, 97–101.

Izenman, A. J. (1986). Bootstrapping Kolmogorov–Smirnov statistics II. *Proc. Comput. Sci. Statist.* **18**, 363–366.

Izenman, A. J., and Sommer, C. J. (1988). Philatelic mixtures and multimodal densities. *J. Am. Statist. Assoc.* **83**, 941–953.

Jacoby, W. G. (1992). PROC IML Statements for Creating a Bootstrap Distribution of OLS Regression Coefficients (assuming random regressors). Technical Report. University of South Carolina.

Jagoe, R. H., and Newman, M. C. (1997). Bootstrap estination of community NOEC values. *Ecotoxicology* **6**, 293–306.

Jain, A. K., Dubes, R. C., and Chen, C. (1987). Bootstrap techniques for error estimation. *IEEE Trans. Pattern Anal. Machine Intell.* **PAMI-9**, 628–633.*

James. G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika* **38**, 324–329.*

James. G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when ratios of population variances are unknown. *Biometrika* **41**, 19–43.*

James. G. S. (1955). Cumulants of a transformed variate. *Biometrika* **42**, 529–531.*

James. G. S. (1958). On moments and cumulants of systems of statistics. *Sankhya* **20**, 1–30.*

James, L. F. (1993). The bootstrap, Bayesian bootstrap and random weighted methods for censored data models. Ph.D. dissertation, State University of New York at Buffalo.

James, L. F. (1997). A study of a class of weighted bootstraps for censored data. *Ann. Statist.* **25**, 1595–1621.

Janas, D. (1991). A smoothed bootstrap estimator for a studentized sample quantile. Preprint. Sonderforschungsbereich 123, Universitat Heidelberg.

Janas, D. (1993). *Bootstrap Procedures for Time Series*. Verlag Shaker, Aachen.*

Janas, D., and Dahlhaus, R. (1994). A frequency domain bootstrap for time series. In *26th Conference on the Interface Between Computer Science and Statistics.* (J. Sall and A. Lehman, editors), Vol. 26, pp. 423–425.

Janssen, P. (1994). Weighted bootstrapping of *U*-statistics. *J. Statist. Plann. Inf.* **38**, 31–42.

Jayasuriya, B. R. (1996). Testing for polynomial regression using nonparametric regression technique. *J. Am. Statist. Assoc.* **91**, 1626–1631.

Jennison, C. (1992). Bootstrap tests and confidence intervals for a hazard ratio when the number of observed failures is small, with applications to group sequential survival studies. In *Computer Science and Statistics: Proceedings of the 22nd Symposium on the Interface* (C. Page and R. LePage, editors), pp. 89–97. Springer-Verlag, New York.

Jensen, J. L. (1992). A modified signed likelihood statistic and saddlepoint approximation. *Biometrika* **79**, 693–703.*

Jensen, J. L. (1995). *Saddlepoint Approximations*. Clarendon Press, Oxford.*

Jensen, R. L., and Kline, G. M. (1994). The resampling cross-validation technique in exercise science: modeling rowing power. *Med. Sci. Sports Exerc.* **26**, 929–933.

Jeong, J., and Maddala, G. S. (1993). A perspective on application of bootstrap methods in econometrics. In *Handbook of Statistics 11: Econometrics* (G. S. Maddala, C. R. Rao, and H. D. Vinod, editors), pp. 573–610. North-Holland, Amsterdam.*

Jeske, D. R., and Marlow, N. A. (1997). Alternative prediction intervals for Pareto proportions. *J. Qual. Technol.* **29**, 317–326.

Jhun, M. (1988). Bootstrapping density estimates. *Commun. Statist. Theory Methods* **17**, 61–78.

Jhun, M. (1990). Bootstrapping k-means clustering. *J. Jpn. Soc. Comput. Statist.* **3**, 1–14.

Jing, B.-Y., and Wood, A. T. A. (1996). Exponential empirical likelihood is not Bartlett correctable. *Ann. Statist.* **24**, 365–369.

Jockel, K.-H. (1986). Finite sample properties and asymptotic efficiency of Monte Carlo tests. *Ann. Statist.* **14**, 336–347.

Jockel, K.-H. (1990). Monte Carlo techniques and hypothesis testing. In *The Frontiers of Statistical Computation, Simulation, and Modeling.* Vol. 1 *Proceedings of the ICOSCO–I Conference, 21–42* (P. R. Nelson, A. Öztürk, E. Dudewicz, and E. C. vander Meulen, editors), American Sciences Press, Syracuse.

Jockel, K.-H., Rothe, G., and Sendler, W. (editors) (1992). Bootstrapping and Related Techniques. *Lecture Notes in Economics and Mathematical Systems*, Vol. 376. Springer-Verlag, Berlin.*

Johns, M. V., Jr. (1988). Importance sampling for bootstrap confidence intervals. *J. Am. Statist. Assoc.* **83**, 709–714.*

Johnson, M. E. (1987). *Multivariate Statistical Simulation*. Wiley, New York.*

Johnson, M. E. (1992). Some modelling and simulation issues related to bootstrapping. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 227–232. Springer-Verlag, Berlin.

Johnson, N. L., and Kotz, S., editors (1982). Bootstrapping. In *Encyclopedia of Statistical Sciences*, Vol. 1, p. 301, Wiley, New York.

Johnson, N. L., and Kotz, S. (1977). *Urn Models*. Wiley, New York.*

Johnsson, T. (1988). Bootstrap multiple test procedures. *J. Appl. Statist.* **15**, 335–339.

Johnstone, I. M., and Velleman, P. F. (1985). Efficient scores, variance decompositions and Monte Carlo swindles. *J. Am. Statist. Assoc.* **80**, 851–862.

Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer-Verlag, New York.*

Jones, G. K. (1988). Sampling errors (computation of). In *Encyclopedia of Statistical Sciences*, Vol. 8, pp. 241–246. Wiley, New York.

Jones, G., Wortberg, M., Kreissig, S. B., Hammock, B. D., and Rocke, D. M. (1996). Application of the bootstrap to calibration experiments. *Anal. Chem.* **68**, 763–770.*

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *J. Am. Statist. Assoc.* **91**, 401–407.

Jones, P. W., Ashour, S. K., and El-Sayed, S. M. (1991). Bootstrap investigation on mixed exponential model using maximum likelihood. In *26th Annual Conference on Statistics, Computer Science and Operations Research*, Vol. 1: *Mathematical Statistics*, pp. 55–78. Cairo University.

Jones, P., Lipson, K., and Phillips, B. (1994). A role for computer-intensive methods in introducing statistical inference. In *Proceedings of the 1st Scientific Meeting—International Association for Statistical Education* (L. Brunelli, and G. Cicchitelli, editors), Vol. 1, pp. 255–264. Universita di Perguria, Perguria.

Joseph, L., and Wolfson, D. B. (1992). Estimation in multi-path change-point problems. *Commun. Statist. Theory and Methods* **21**, 897–913.

Joseph, L., Wolfson, D. B., du Berger, R., and Lyle, R. M. (1996). Change-point analysis of a randomized trial on the effects of calcium supplementation on blood pressure. In *Bayesian Biostatistics* (D. A. Berry and D. K. Stangl, editors), pp. 617–649. Marcel Dekker, New York.

Journel, A. G. (1994). Resampling from stochastic simulations (with discussion). *Environ. Ecol. Statist.* **1**, 63–91.

Junghard, O. (1990). Linear shrinkage in traffic accident models and their estimation by cross validation and bootstrap methods. Linkoping Studies in Science and Technology, Thesis No. 205, Linkoping.

Jupp, P. E., and Mardia, K. V. (1989). Theory of directional statistics, 1975–1988. *Int. Statist. Rev.* **57**, 261–294.

Kabaila, P. (1993a). On bootstrap predictive inference for autoregressive process. *J. Time Ser. Anal.* **14**, 473–484.*

Kabaila, P. (1993b). Some properties of profile bootstrap confidence intervals. *Austral. J. Statist.* **35**, 205–214.

Kadiyala, K. R., and Qberhelman, D. (1990). Estimation of standard errors of empirical Bayes estimators in CAPM-type models. *Commun. Statist. Simul. Comput.* **19**, 189–206.

Kafadar, K. (1994). An application of nonlinear regression in R & D: A case study from the electronics industry. *Technometrics* **36**, 237–248.

Kaigh, W. D., and Cheng, C. (1991). Subsampling quantile estimator standard errors with applications. *Commun. Statist. Theory Methods* **20**, 977.

Kaigh, W. D., and Lachenbruch, P. (1982). A generalized quantile estimator. *Commun. Statist. Theory Methods* **11**, 2217–2238.

Kalbfleish, J. D., and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

Kanal, L. (1974). Patterns in pattern recognition: 1968–1974. *IEEE Trans. Inform. Theory* **2**, 472–479.*

Kane, V. E. (1986). Process capability indices. *J. Qual. Technol.* **24**, 41–52.*

Kang, S.-B., and Cho, Y.-S. (1997). Estimation of the parameters on a Pareto distribution by jackknife and bootstrap. *J. Inform. Optim. Sci.* **18**, 289–300.

Kaplan, E. L., and Meier, P. (1958). Nonparametric estimation from incomplete samples. *J. Am. Statist. Assoc.* **53**, 457–481.*

Kapoyannis, A. S., Ktorides, C. N., and Panagiotou, A. D. (1997). An extension of the statistical bootstrap model ot include strangeness. *J. Phys. G. Nucl. Particle Phys. Strangeness in Quark Matter 1997* **23**, 1921–1932.

Kapoyannis, A. S., Ktorides, C. N., and Panagiotou, A. D. (1998). An extension of the statistical bootstrap model to include strangeness. Implications on particle ratios. *Phys. Rev. D* **58**, 1–17.

Karrison, T. (1990). Bootstrapping censored data with covariates. *J. Statist. Comput. Simul.* **36**, 195–207.

Katz, A. S., Katz, S., and Lowe, N. (1994). Fundamentals of the bootstrap based analysis of neural network's accuracy. In *Proceedings of the World Congress on Neural Networks*, Vol. 3, pp. 673–678.

Kaufman, E. (1988). Asymptotic expansion of the distribution function of the sample quantile prepivoted by the bootstrap. Diplomarbeit thesis, University of Siegen, Siegen (in German).

Kaufman, L., and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Kaufman, S. (1993). A bootstrap variance estimator for the Schools and Staffing Survey. In *ASA Proceedings of the Survey Research Section*, pp. 675–680.

Kaufman, S. (1998). A bootstrap variance estimator for systematic PPS sampling. Unpublished paper.*

Kawano, H., and Higuchi, T. (1995). The bootstrap method in space physics: error estimation for the minimum variance analysis. *Geophys. Res. Lett.* **22**, 307–310.

Kay, J., and Chan, K. (1992). Bootstrapping blurred and noisy data. *Proceedings 10th Symposium on Computational Statistics* (Y. Dodge and J. Whittaker, editors), Vol. 2, pp. 287–291. Physica, Vienna.

Keating, J. P., and Tripathi, R. C. (1985). Estimation of percentiles. In *Encyclopedia of Statistical Sciences*, Vol. 6, pp. 668–674. Wiley, New York.

Kemp, A. W. (1997). Book review of *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd ed., by B. F. J. Manly. *Biometrics* **53**, 1560–1561.

Kendall, D. G., and Kendall, W. S. (1980). Alignments in two-dimensional random sets and points. *Adv. Appl. Probab.* **12**, 380–424.

Kim, D. (1993). Nonparametric kernel regression function estimation with bootstrap method. *J. Korean Statist. Soc.* **22**, 361–368.

Kim, H. T., and Truong, Y. K. (1998). Nonparametric regression estimates with censored data: local linear smoothers and their applications. *Biometrics* **54**, 1434–1444.

Kim, J.-H. (1990). Conditional bootstrap methods for censored data. Ph.D. Thesis, Department of Statistics, Florida State University.

Kim, Y. B., Haddock, J., and Willemain, T. R. (1993). The binary bootstrap: Inference with autocorrelated binary data. *Commun. Statist. Simul. Comput.* **22**, 205–216.

Kimber, A. (1994). Book Review of *An Introduction to the Bootstrap* by B. Efron and R. Tibshirani. *Statistician* **43**, 600.

Kinateder, J. G. (1992). An invariance principle applicable to the bootstrap. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 157–181. Wiley, New York.

Kindermann, J., Paass, G., and Weber, F. (1995). Query construction for neural networks using the bootstrap. In *ICANN '95* (F. Fogelman-Soulie and P. Gallinari, editors), Vol. **2**, pp. 135–140.

Kinsella, A. (1987). The 'exact' bootstrap approach to confidence intervals for the relative difference statistic. *Statistician* **36**, 345–347.

Kipnis, V. (1992). Bootstrap assessment of prediction in exploratory regression analysis. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 363–387. Wiley, New York.

Kish, L., and Frankel, M. R. (1970). Balanced repeated replication for standard errors. *J. Am. Statist. Assoc.* **65**, 1071–1094.

Kish, L., and Frankel, M. R. (1974). Inference from complex samples (with discussion) *J. R. Statist. Soc. B* **36**, 1–37.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084–2102.

Klenitsky, D. V., and Kuvshinov, V. I. (1996). Local fluctuations of multiplicity in the statistical-bootstrap model. *Phys. At. Nucl.* **59**, 129–134.

Klenk, A, and Stute, W. (1987). Bootstrapping of L-estimates. *Statist. Dec.* **5**, 77–87.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (1998). *Loss Models: From Data to Decisions*. Wiley, New York.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Ann. Statist.* **17**, 1168–1175.*

Knight, K. (1997). Bootstrapping sample quantiles in non-regular cases. *Statist. Probab. Lett.* **37**, 259–267.

Knoke, J. D. (1986). The robust estimation of classification error rates. *Comput. Math. Applic.* **12A**, 253–260.

Knox, R. G., and Peet, R. K. (1989). Bootstrapped ordination: A method for estimating sampling effects in indirect gradient analysis. *Vegetatio* **80**, 153–165.

Kocherlakota, S., Kocherlakota, K., and Kirmani, S. N. U. A. (1992). Process capability indices under non-normality. *Int. J. Math. Statist.* **1**, 175–209.*

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy: Assessment and model selection. Stanford University Technical Report, Department of Computer Sciences.

Kolen, M. J., and Brennan, R. L. (1987). Analytic smoothing for equipercentile equating under the common item nonequivalent population design. *Psychometrika* **52**, 43–59.

Koltchinskii, V. I. (1997). *M*-estimation convexity and quantiles. *Ann. Statist.* **25**, 435–477.

Kong, F., and Zhang, M. (1994). The Edgeworth expansion and the bootstrap approximation for a studentized *U*-statistic. In *5th Japan–China Symposium on Statistics* (M. Ichimura, S. Mao, and G. Fan, editors), Vol. 5, pp. 124–126.

Konishi, S. (1991). Normalizing transformations and bootstrap confidence intervals. *Ann. Statist.* **19**, 2209–2225.

Konishi, S., and Honda, M. (1990). Comparison of procedures for estimation of error rates in discriminant analysis under nonnormal populations. *J. Statist. Comput. Simul.* **36**, 105–116.

Konold, C. (1994). Understanding probability and statistical inference through resampling. In *Proceedings of the 1st Scientific Meeting-International Association for Statistical Education* (L. Brunelli, and G. Cicchitelli, editors), Vol. 1, pp. 199–212, Universita di Perugia, Perugia.

Kononenko, I. V., and Derevyanchenko, B. I. (1995). Analysis of an algorithm for prediction of stationary random processes using the bootstrap. *Automatika* **2**, 88–92.

Kotz, S., and Johnson, N. L. (editors) (1992). *Breakthroughs in Statistics: Methodology and Distribution*, Vol. 2, Springer-Verlag, New York.*

Kotz, S., and Johnson, N. L. (1993). *Process Capability Indices*. Chapman & Hall, London.*

Kotz, S., and Johnson, N. L. (editors) (1997). *Breakthroughs in Statistics*, Vol. 3, Springer-Verlag, New York.*

Kotz, S., Johnson, N. L., and Read, C. B., (editors) (1982). Bootstrapping. In *Encyclopedia of Statistical Sciences*, Vol. 1, p. 301. Wiley, New York.*

Kotz, S., Johnson, N. L., and Read, C. B. (editors) (1983). Gauss–Markov Theorem. In *Encyclopedia of Statistical Sciences*, Vol. 3, pp. 314–316. Wiley, New York.*

Koul, H., and Lahiri, S. N. (1994). On bootstrapping *M*-estimated residual processes in multiple linear regression models. *J. Mult. Anal.* **49**, 255–265.

Kovar, J. G. (1985). Variance estimation of nonlinear statistics in stratified samples. Methodology Branch Working Paper # 85-052E, Statistics Canada.*

Kovar, J. G. (1987). Variance estimation of medians in stratified samples. Methodology Branch Working Paper # 87-004E, Statistics Canada.*

Kovar, J. G., Rao, J. N. K., and Wu, C. F. J. (1988). Bootstrap and other methods to measure errors in survey estimates. *Can. J. Statist.* **16** (Suppl.), 25–45.*

Kreiss, J.-P. (1992). Bootstrap procedures for AR(8)—processes. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 107–113. Springer-Verlag, Berlin.

Kreiss, J.-P., and Franke, J. (1992). Bootstrapping stationary autoregressive moving average models. *J. Time Ser. Anal.* **13**, 297–317.

Krewski, D., and Rao, J. N. K. (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.* **9**, 1010–1019.

Krewski, D., Smythe, R. T., Dewanji, A., and Szyszkowicz, M. (1991a). Bootstrapping an empirical Bayes estimator of the distribution of historical controls in carcinogen bioassay. Unpublished manuscript.

Krewski, D., Smythe, R. T., Fung, K. Y., and Burnett, R. (1991b). Conditional and unconditional tests with historical controls. *Can. J. Statist.* **19**, 407–423.

Kuk, A. Y. C. (1987). Bootstrap estimators of variance under sampling with probability proportional to aggregate size. *J. Statist. Comput. Simul.* **28**, 303–311.*

Kuk, A. Y. C. (1989). Double bootstrap estimation of variance under systematic sampling with probability proportional to size. *J. Statist. Comput. Simul.* **31**, 73–82.*

Kulperger, P. J., and Prakasa Rao, B. L. S. (1989). Bootstrapping a finite state Markov chain. *Sankhya A* **51**, 178–191.*

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217–1241.*

Lachenbruch, P. A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics* **23**, 639–645.*

Lachenbruch, P. A. (1975). *Discriminant Analysis*. Hafner, New York.*

Lachenbruch, P. A., and Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics* **10**, 1–11.*

Laeuter, H. (1985). An efficient estimator of error rate in discriminant analysis. *Statistics* **16**, 107–119.

Lahiri, S. N. (1991). Second order optimality of stationary bootstrap. *Statist. Probab. Lett.* **14**, 335–341.*

Lahiri, S. N. (1992a). On bootstrapping *M*-estimators. *Sankhya A* **54**, 157–170.*

Lahiri, S. N. (1992b). Edgeworth correction by moving block bootstrap for stationary and nonstationary data. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 183–214. Wiley, New York.*

Lahiri, S. N. (1992c). Bootstrapping *M*-estimators of a multiple linear regression parameter. *Ann. Statist.* **20**, 1548–1570.*

Lahiri, S. N. (1993a). Bootstrapping the studentized sample mean of lattice variables. *J. Multi. Anal.* **45**, 247–256.

Lahiri, S. N. (1993b). On the moving block bootstrap under long range dependence. *Statist. Probab. Lett.* **18**, 405–413.

Lahiri, S. N. (1994a). On second order correctness of Efron's bootstrap without Cramer-type conditions in linear regression models. *Math. Meth. Statist.* **3**, 130–148.

Lahiri, S. N. (1994b). Rates of bootstrap approximation for the mean of lattice variables. *Sankhya A* **56**, 77–89.

Lahiri, S. N. (1994c). Two term Edgeworth expansion and bootstrap approximation for multivariate studentized *M*-estimators. *Sankhya A* **56**, 201–226.

Lahiri, S. N. (1995). On asymptotic behavior of the moving block bootstrap for normalized sums of heavy-tailed random variables. *Ann. Statist.* **23**, 1331–1349.*

Lahiri, S. N. (1996). On Edgworth expansions and the moving block bootstrap for studentized *M*-estimators in multiple linear regression models. *J. Multi. Anal.* **56**, 42–59.

Lahiri, S. N. (1997). Bootstrapping weighted empirical processes that do not converge weakly. *Statist. Probab. Lett.* **37**, 295–302.

Lahiri, S. N., and Koul, H. (1994). On bootstrapping *M*-estimated residual processes in multiple linear regression models. *J. Mult. Anal.* **49**, 255–265.

Lai, T. L., and Wang, J. Q. Z. (1993). Edgeworth expansions for symmetric statistics with applications to bootstrap methods. *Statist. Sin.* **3**, 517–542.

Laird, N. M., and Louis, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples (with discussion). *J. Am. Statist. Assoc.* **82**, 739–757.

Lake, J. A. (1995). Calculating the probability of multitaxon evolutionary trees: Bootstrappers gambit. *Proc. Natl. Acad. Sci.* **92**, 9662–9666.

Lamb, R. H., Boos, D. D., and Brownie, C. (1996). Testing for effects on variance in experiments with factorial treatment structure and nested errors. *Technometrics* **38**, 170–177.

Lambert, D., and Tierney, L. (1997). Nonparametric maximum likelihood estimation from samples with irrelevant data and verification bias. *J. Am. Statist. Assoc.* **92**, 937–944.

LaMotte, L. R. (1978). Bayes linear estimators. *Technometrics* **20**, 281–290.*

Lancaster, T. (1997). Exact structural inference in optimal job-search models. *J. Bus. Econ. Statist.* **15**, 165–179.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the *t* distribution. *J. Am. Statist. Assoc.* **84**, 881–905.

Lanyon, S. M. (1987). Jackknifing and bootstrapping: Important "new" statistical techniques for ornithologists. *Auk* **104**, 144–146.*

Lavori, P. W., Dawson, R., and Shera, D. (1995). A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine*, **14**, 1913–1925.

Lawless, J. F. (1988). Reliability (nonparametric methods in). In *Encyclopedia of Statistical Sciences*, Vol. 8, pp. 20–24. Wiley, New York.*

Leadbetter, M. R., Lindgren, G., and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.*

Leal, S. M., and Ott, J. (1993). A bootstrap approach to estimating power for linkage heterogeneity. *Genet. Epidemiol.* **10**, 465–470.*

LeBlanc, M., and Crowley, J. (1993). Survival trees by goodness of split. *J. Am. Statist. Assoc.* **88**, 477–485.*

LeBlanc, M., and Tibshirani, R. (1996). Combining estimates in regression and classification. *J. Am. Statist. Assoc.* **91**, 1641–1650.

Lebreton, C. M., and Visscher, P. M. (1998). Empirical nonparametric bootstrap strategies in quantitative trait loci mapping: Conditioning on the genetic model. *Genetics* **148**, 525–535.

Lee, A. J. (1985). On estimating the variance of a *U*-statistic. *Commun. Statist. Theory Methods* **14**, 289–301.

Lee, K. W. (1990). Bootstrapping logistic regression models with random regressors. *Commun. Statist. Theory Methods* **19**, 2527–2539.

Lee, S. M. S. (1994). Optimal choice between parametric and nonparametric estimates. *Math. Proc. Cambridge Philos. Soc.* **115**, 335–363.

Lee, S. M. S. (1998). On a class of *m*-out-of-*n* bootstrap confidence intervals. *Research Report 195*, Department of Statistics, University of Hong Kong.

Lee, S. M. S., and Young, G. A. (1994a). Asymptotic iterated bootstrap confidence intervals. In *26th Conference on the Interface Between Computer Science and Statistics* (J. Sall and A. Lehman, editors), Vol. 26, pp. 464–471, Springer-Verlag, New York.

Lee, S. M. S., and Young, G. A. (1994b). Practical higher-order smoothing of the bootstrap. *Statist. Sin.* **4**, 445–460.

Lee, S. M. S., and Young, G. A. (1995). Asymptotic iterated bootstrap confidence intervals. *Ann. Statist.* **23**, 1301–1330.

Lee, S. M. S., and Young, G. A. (1997). Asymptotic and resampling methods. In *28th Symposium on the Interface Between Computer Science and Statistics* (L. Billard, and N. I. Fisher, editors), Vol. 28, pp. 221–227, Springer-Verlag, New York.

Leger, C., and Cleroux, R. (1992). Nonparametric age replacement: Bootstrap confidence intervals for the optimal cost. *Oper. Res.* **40**, 1062–1073.

Leger, C., and Larocque, D. (1994). Bootstrap estimates of the power of a rank test in a randomized block design. *Statist. Sin.* **4**, 423–443.

Leger, C., Politis, D. N., and Romano, J. P. (1992). Bootstrap technology and applications. *Technometrics* **34**, 378–398.*

Leger, C., and Romano, J. P. (1990a). Bootstrap adaptive estimation. *Can. J. Statist.* **18**, 297–314.

Leger, C., and Romano, J. P. (1990b). Bootstrap choice of tuning parameters. *Ann. Inst. Statist. Math.* **42**, 709–735.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York.*

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer-Verlag, New York.*

Lehmann, E. L., and Casella, G. E. (1998). *Theory of Point Estimation*, 2nd ed., Springer-Verlag, New York.*

Lehtonen, R., and Pakkinen, E. J. (1995). *Practical Methods for the Design and Analysis of Complex Surveys*. Wiley, Chichester.

Lele, S. (1991a). Jackknifing linear estimating equations: Asymptotic theory and applications in stochastic processes. *J. R. Statist. Soc. B* **53**, 253–267.*

Lele, S. (1991b). Resampling using estimating equations. In *Estimating Functions* (V. Godambe, editor), pp. 295–304. Clarendon Press, Oxford.*

LePage, R. (1992). Bootstrapping signs. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 215–224. Wiley, New York.*

LePage, R., and Billard, L. (editors) (1992). *Exploring the Limit of Bootstrap*. Wiley, New York.*

LePage, R., Podgórski, K., Ryznar, M., and White, A. (1998). Bootstrapping Signs and Permutations for Regression with Heavy-Tailed Errors: a Robust Resampling. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. R. J. Adler, R. E. Feldman, and M. S. Taqqu, Editors, pp. 339–358. Birkhäuser, Boston.*

Li, H., and Maddala, G. S. (1996). Bootstrapping time series models (with discussion). *Econ. Rev.* **15**, 115–195.*

Li, G., Tiwari, R. C., and Wells, M. T. (1996). Quantile comparison functions in two-sample problems, with application to comparisons of diagnostic markers. *J. Am. Statist. Assoc.* **91**, 689–698.

Linder, E., and Babu, G. J. (1994). Bootstrapping the linear functional relationship with known error variance ratio. *Scand. J. Statist.* **21**, 21–39.

Lindsay, B. G., and Li, B. (1997). On second-order optimality of the observed Fisher information. *Ann. Statist.* **25**, 2172–2199.

Linhart, H., and Zucchini, W. (1986). *Model Selection*. Wiley, New York.*

Linnet, K. (1989). Assessing diagnostic tests by a strictly proper scoring rule. *Statist. Med.* **8**, 609–618.

Linssen, H. N., and Nanens, P. J. A. (1983). Estimation of the radius of a circle when the coordinates of a number of points on its circumference are observed: An example of bootstrapping. *Statist. Probab. Lett.* **1**, 307–311.

Little, R. J. A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Appl. Statist.* **37**, 23–38.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.*

Liu, J. (1992). Inference from Stratified Samples: Application of Edgeworth Expansions. Ph.D. thesis, Carleton University.

Liu, J. S., and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Am. Statist. Assoc.* **93**, 1032–1044.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16**, 1696–1708.*

Liu, R. Y., and Singh, K. (1987). On a partial correction by the bootstrap. *Ann. Statist.* **15**, 1713–1718.*

Liu, R. Y., and Singh, K. (1992a). Efficiency and robustness in sampling. *Ann. Statist.* **20**, 370–384.

Liu, R. Y., and Singh, K. (1992b). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 225–248. Wiley, New York.*

Liu, R. Y., and Singh, K. (1995). Using i.i.d. bootstrap for general non-i.i.d. models. *J. Statist. Plann. Inf.* **43**, 67–76.*

Liu, R. Y., and Singh, K. (1997). Notions of limiting p values based on data depth and the bootstrap. *J. Am. Statist. Assoc.* **92**, 266–277.

Liu, R. Y., Singh, K., and Lo, S. H. (1989). On a representation related to bootstrap. *Sankhya A* **51**, 168–177.

Liu, R. Y., and Tang, J. (1996). Control charts for dependent and independent measurements based on bootstrap methods. *J. Am. Statist. Assoc.* **91**, 1694–1700.*

Liu, Z. J. (1991). Bootstrapping one way analysis of Rao's quadratic entropy. *Comm. Statist.* **20**, 1683–1702.

Liu, Z. J., and Rao, C. R. (1995). Asymptotic distribution of statistics based on quadratic entropy and bootstrapping. *J. Statist. Plann. Inf.* **43**, 1–18.

Lloyd, C. L. (1998). Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *J. Am. Statist. Assoc.* **93**, 1356–1364.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12**, 351–357.

Lo, A. Y. (1987). A large sample sample study of the Bayesian bootstrap. *Ann. Statist.* **15**, 360–375.*

Lo, A. Y. (1988). A Bayesian bootstrap for a finite population. *Ann. Statist.* **16**, 1684–1695.*

Lo, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhya A* **53**, 320–333.

Lo, A. Y. (1993a). A Bayesian bootstrap for censored data. *Ann. Statist.* **21**, 100–123.*

Lo, A. Y. (1993b). A Bayesian bootstrap for weighted sampling. *Ann. Statist.* **21**, 2138–2148.

Lo, S.-H. (1998). General coverage problems with applications and bootstrap method in survival analysis and reliability theory. Columbia University Technical Report.

Lo, S.-H., and Singh, K. (1986). The product limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory* **71**, 455–465.

Lo, S.-H., and Wang, J. L. (1989). I.I.D. representations for bivariate product limit estimators and the bootstrap versions. *J. Multivar. Anal.* **28**, 211–226.

Lodder, R. A., Selby, M., and Hieftje, G. M. (1987). Detection of capsule tampering by near infra-red reflectance analysis. *Anal. Chem.* **59**, 1921–1930.

Loh, W.-Y. (1984). Estimating an endpoint of a distribution with resampling methods. *Ann. Statist.* **12**, 1543–1550.

Loh, W.-Y. (1985). A new method for testing separate families of hypotheses. *J. Am. Statist. Assoc.* **80**, 362–368.

Loh, W.-Y. (1987). Calibrating confidence coefficients. *J. Am. Statist. Assoc.* **82**, 155–162.*

Loh, W.-Y. (1991). Bootstrap calibration for confidence interval construction and selection. *Statist. Sin.* **1**, 479–495.

Lohse, K. (1987). Consistency of the bootstrap. *Statist. Dec.* **5**, 353–366.

Lokki, H., and Saurola, P. (1987). Bootstrap methods for the two-sample location and scatter problems. *Acta Ornithol.* **23**, 133–147.

Lombard, F. (1986). The change-point problem for angular data. *Technometrics* **28**, 391–397.

Loughin, T. M., and Koehler, K. (1993). Bootstrapping in proportional hazards models with fixed explanatory variables. Unpublished manuscript.

Loughin, T. M., and Noble, W. (1997). A permutation test for effects in an unreplicated factorial design. *Technometrics* **39**, 180–190.

Lovie, A. D., and Lovie, P. (1986). The flat maximum effect and linear scoring models for prediction. *J. For.* **5**, 159–168.

Low, L. Y. (1988). Resampling procedures. In *Encyclopedia of Statistical Sciences*, Vol. 8, pp. 90–93. Wiley, New York.

Lu, H. H. S., Wells, M. T., and Tiwari, R. C. (1994). Inference for shift functions in the two-sample problem with right-censored data: With applications. *J. Am. Statist. Assoc.* **89**, 1017–1026.

Lu, J.-C., Park, J., and Yang, Q. (1997). Statistical inference of a time-to-failure distribution derived from linear degradation data. *Technometrics* **39**, 391–400.

Lu, M.-C., and Chang, D. S. (1997). Bootstrap prediction intervals for the Birnbaum–Saunders distribution. *Microelectron. Reliab.* **37**, 1213–1216.

Lu, R., and Yang, C.-H. (1994). Resampling schemes for estimating the similarity measures on location model. *Commun. Statist. Simul. Comput.* **23**, 973–996.

Ludbrook, J. (1995). Issues in biomedical statistics; comparing means by computer-intensive methods. *Aust. N. Z. J. Surg.* **65**, 812–819.

Lunneborg, C., E. (1985). Estimating the correlation coefficient: The bootstrap approach. *Psych. Bull.* **98**, 209–215.*

Lunneborg, C. E. (1987). Bootstrap applications for the behavioral sciences. *Psychometrika* **52**, 477–478.*

Lunneborg, C. E. (1994). Book review of *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* by P. Westfall and S. S. Young. *Statist. Comp.* **4**, 219–220.

Lunneborg, C. E., and Tousignant, J. P. (1985). Efron's bootstrap with application to the repeated measures design. *Multivar. Behav. Res.* **20**, 161–178.

Lutz, M. W., Kenakin, T. P., Corsi, M., Menius, J. A, Krishnamoorthy, C., Rimele, T., and Morgan, P. H. (1995). Use of resampling techniques to estimate the variance of parameters in pharmacological assays when experimental protocols preclude independent replication: An example using Schild regressions. *J. Pharmacol. Toxicol. Methods* **34**, 37–46.

Magnussen, S., and Burgess, D. (1997). Stochastic resampling techniques for quantifying error propagations in forest field experiments. *Can. J. For. Res.* **27**, 630–637.

Maindonald, J. H. (1984). *Statistical Computation*. Wiley, New York.

Maitra, R. (1997). Estimating precision in functional images. *J. Comp. Graph. Statist.* **6**, 285–299.

Maiwald, D., and Bohme, J. F. (1994). Multiple testing for seismic data using bootstrap. In *Proceedings of the 1994 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 6, pp. 89–92.

Mak, T. K., Li, W. K., and Kuk, A. Y. C. (1986). The use of surrogate variables in binary regression models. *J. Statist. Comput. Simul.* **24**, 245–254.

Makinodan, T., Albright, J. W., Peter, C. P., Good, P. I., and Heidrick, M. L. (1976). Reduced humoral immune activity in long-lived mice. *Immunology* **31**, 400–408.*

Mallows, C. L. (1983). Robust methods. Chapter 3 in *Statistical Data Analysis* (R. Gnanadesikan, editor). American Mathematical Society, Providence.

Mallows, C. L., and Tukey, J. W. (1982). An overview of techniques of data analysis, Emphasizing its exploratory aspects. In *Some Recent Advances in Statistics* (J. T. de Oliveira and B. Epstein, editors), pp. 111–172. Academic Press, London.

Mammen, E. (1989a). Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *Ann. Statist.* **17**, 382–400.

Mammen, E. (1989b). Bootstrap and wild bootstrap for high-dimensional linear models. Preprint. Sonderforschungsbereich 123, Universitat Heidelberg.

Mammen, E. (1990). Higher order accuracy of bootstrap for smooth functionals. Preprint. Sonderforschungsbereich 123, Universitat Heidelberg.

Mammen, E. (1992a). Bootstrap, wild bootstrap, and asymptotic normality. Preprint. Sonderforschungsbereich 123, Universitat Heidelberg.

Mammen, E. (1992b). When Does the Bootstrap Work? Asymptotic Results and Simulations. *Lecture Notes in Statistics*, Vol. 77. Springer-Verlag, Heidelberg.*

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21**, 255–285.*

Mammen, E., Marron, J. S., and Fisher, N. I. (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Probab. Th. Rel. Fields* **91**. 115–132.

Manly, B. F. J. (1991). *Randomization and Monte Carlo Methods in Biology*. Chapman & Hall, London.*

Manly, B. F. J. (1992). Bootstrapping for determining sample sizes in biological studies. *J. Exp. Mar. Biol. Ecol.* **158**, 189–196.

Manly, B. F. J. (1993). A review of computer-intensive multivariate methods in ecology. In *Multivariate Environmental Statistics* (G. P. Patil, and C. R. Rao, editors), pp. 307–346. Elsevier Science Publishers, Amsterdam.*

Manly, B. F. J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd ed. Chapman & Hall, London.*

Mao, X. (1996). Splatting of non-rectilinear volumes through stochastic resampling. *IEEE Trans. Vis. Comput. Graphics* **2**, 156–170.

Mapleson, W. W. (1986). The use of GLIM and the bootstrap in assessing a clinical trial of two drugs. *Statist. Med.* **5**, 363–374.*

Mardia, K.V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, London.*

Maritz, J. S. (1979). A note on exact robust confidence intervals for location. *Biometrika* **66**, 163–166.*

Maritz, J. S., and Jarrett, R. G. (1978). A note on estimating the variance of the sample median. *J. Am. Statist. Assoc.* **73**, 194–196.*

Maritz, J. S., and Lwin, T. (1989). *Empirical Bayes Methods*, 2nd ed. Chapman & Hall, London.*

Markus, M. T. (1994). Bootstrap confidence regions for homogeneity analysis. The influence of rotation on coverage percentages. *Proc. Comput. Statist. Symp.* **11**, 337–342.

Markus, M. T., and Visser, R. A. (1990). Bootstrap methods for generating confidence regions in HOMALS; Balancing sample size and number of trials. FSW/RUL, RR-90–02, Department of Behavioral Computer Science, University of Leiden, Leiden.

Markus, M. T., and Visser, R. A. (1992). Applying the bootstrap to generate confidence regions in multiple correspondence analysis. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 71–75. Springer-Verlag, Berlin.

Marron, J. S. (1992). Bootstrap bandwidth selection. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 249–262. Wiley, New York.*

Martin, M. A. (1989). On the bootstrap and confidence intervals. Unpublished Ph.D. thesis, Australian National University.

Martin, M. A. (1990a). On the bootstrap iteration for coverage correction in confidence intervals. *J. Am. Statist. Assoc.* **85**, 1105–1118.*

Martin, M. A. (1990b). On using the jackknife to estimate quantile variance. *Can. J. Statist.* **18**, 149–153.

Martin, M. A. (1994a). Book review of *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. *Biometrics* **50**, 1226–1227.

Martin, M. A. (1994b). Book review of *An Introduction to Bootstrap* by B. Efron and R. Tibshirani. *Chance* **7**, 58–60.

Martin. R. D. (1980). Robust estimation of autoregressive models. In *Reports on Directions in Time Series* (D. R. Brillinger and G. C. Tiao, editors), pp. 228–262. Institute of Mathematical Statistics, Hayward.*

Martz, H. F., and Duran, B. S. (1985). A comparison of three methods for calculating lower confidence limits on system reliability using binomial component data. *IEEE Trans. Reliab.* **R-34**, 113–120.

Mason, D. M., and Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.* **20**, 1611–1624.

Matheron, G. (1975). *Random Sets and Integral Geometry*. Wiley, New York.*

Mattei, G., Migani, S., and Rosa, R. (1997). Statistical resampling for accuracy estimate in Monte Carlo renormalization group. *Phys. Lett. A* **237**, 33–36.

Mazurkiewicz, M. (1995). The little bootstrap method for autoregressive model selection. *Badania Operacyjne I Decyzje* **2**, 39–53.

McCarthy, P. J. (1969). Pseudo-replication: Half-samples. *Int. Statist. Rev.* **37**, 239–263.*

McCarthy, P. J., and Snowden, C. B. (1985). The bootstrap and finite population sampling. In *Vital and Health Statistics (Ser. 2 No. 95) Public Health Service Publication* **85-1369**, U.S. Government Printing Office, Washington, D.C.*

McCullough, B. D. (1994). Bootstrapping forecast intervals: An application to AR(p) models. *J. Forecast.* **13**, 51–66.*

McCullough, B. D., and Vinod, H. D. (1998). Implementing the double bootstrap. *Comput. Econ.* **12**, 79–95.*

McDonald, J. A. (1982). Interactive graphics for data analysis. ORION Technical Report 011, Department of Statistics, Stanford University.*

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**. 234–245.*

McKean, J. W., and Schrader, R. M. (1984). A comparison of methods for studentizing the sample quantile. *Commun. Statist. Simul. Comput.* **13**, 751–773.

McLachlan, G. J. (1976). The bias of the apparent error rate in discriminant analysis. *Biometrika* **63**, 239–244.*

McLachlan, G. J. (1980). The efficiency of Efron's bootstrap approach applied to error rate estimation in discriminant analysis. *J. Statist. Comput. Simul.* **11**, 273–279.*

McLachlan, G. J. (1986). Assessing the performance of an allocation rule. *Comput. Math. Applic.* **12A**, 261–272.*

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a mixture. *Appl. Statist.* **36**, 318–324.*

McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, New York.*

McLachlan, G. J., and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York.*

McLachlan, G. J., and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.*

McLachlan, G. J., and Peel, D. (1997). On a Resampling Approach to Choosing the Number of Components in Normal Mixture Models. In *28th Symposium on the Interface Between Computer Science and Statistics* (L. Billard, and N. I. Fisher, editors), Vol. 28, pp. 260–266. Springer-Verlag, New York.

McLeod, A. I. (1988). Simple random sampling. In *Encyclopedia of Statistical Sciences*, **8**, 478–480.

McPeek, M. A., and Kalisz, S. (1993). Population sampling and bootstrapping in complex designs: Demographic analysis. *Ecol. Exper.* 232–252.

McQuarrie, A. D. R., and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*. World Scientific Publishing, Singapore.*

Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley, New York.*

Mehlman, D. W., Shepard, U. L., and Kelt, D. A. (1995). Bootstrapping principal components—a comment (with discussion). *Ecology* **76**, 640–645.

Meneghini, F. (1985). An application of bootstrap for spatial point patterns (in Italian). *R. St. A. Italy.* 73–82.

Meyer, J.S., Ingersoll, C.G., McDonald, L.L., and Boyce, M.S. (1986). Estimating uncertainty in population growth rates: Jackknife vs. bootstrap techniques. *Ecology* **67**, 1156–1166.

Mick, R., and Ratain, M. J. (1994). Bootstrap validation of pharmacodynamic models defined via stepwise linear regression. *Clin. Pharmacol. Ther.* **56**, 217–222.

Mignani, S., and Rosa, R. (1995). The moving block bootstrap to access the accuracy of statistical estimates in Ising model simulations. *Comput. Phys. Commun.* **92**, 203–213.

Mikheenko, S., Erofeeva, S., and Kosako, T. (1994). Reconstruction of dose distribution by the bootstrap method using limited measured data. *Radioisotopes* **43**, 595–604.

Milan, L., and Whittaker, J. (1995). Application of the parametric bootstrap to models that incorporate a singular value decomposition. *Appl. Statist.* **44**, 31–49.

Miller, A. J. (1990). *Subset Selection in Regression*. Chapman & Hall, London.*

Miller, R. G., Jr. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **39**, 1594–1605.

Miller, R. G., Jr. (1974) The jackknife—A review. *Biometrika* **61**, 1–17.*

Miller, R. G., Jr. (1981a). *Survival Analysis*. Wiley, New York.*

Miller, R. G., Jr. (1981b). *Simultaneous Statistical Inference*, 2nd ed. Springer-Verlag, New York.*

Miller, R. G., Jr. (1986). *Beyond ANOVA, Basics of Applied Statistics*. Wiley, New York.*

Miller, R. G., Jr. (1997). *Beyond ANOVA, Basics of Applied Statistics*, 2nd ed. Chapman & Hall, New York.

Milliken, G. A., and Johnson, D. E. (1984). *Analysis of Messy Data,* Vol. 1: *Designed Experiments*. Wadsworth, Belmont.*

Milliken, G. A., and Johnson, D. E. (1989). *Analysis of Messy Data,* Vol. 2*: Nonreplicated Experiments*. Van Nostrand Reinhold, New York.*

Mitani, Y., Hamamoto, Y., and Tomita, S. (1995). Use of bootstrap samples in designing artificial neural network classifiers. In *IEEE International Conference on Neural Networks Proceedings*, Vol. 4, pp. 2103–2106.

Miyakawa, M. (1991). Resampling plan using orthogonal array and its application to influence analysis. *Rep. Statist. Appl. Res., Union of Jpn. Sci. Eng.* **38/2**, 1–10.

Moeher, M. (1987). On the estimation of the expected actual error rate in sample-based multinomial classification. *Statistics* **18**, 599–612.

Monti, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika* **84**, 395–405.

Montvay, I. (1996). Statistics and internal quantum numbers in the statistical bootstrap approach. Hungarian Academy of Sciences, Central Research Institute for Physics, Budapest.

Mooijart, A. (1985). Factor analysis for non-normal variables. *Psychometrika* **50**, 323–342.

Mooney, C. Z. (1996). Bootstrap statistical inference: Examples and evaluations for political science. *Am. J. Political Sci.* **40**, 570–602.*

Mooney, C. Z. (1997). *Monte Carlo Simulation. Quantitative Applications in the Social Sciences*, Vol. 116. Sage Publications, Newberry Park.*

Mooney, C. Z., and Duval, R. D. (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference. Quantitative Applications in the Social Sciences*, Vol. 95. Sage Publications, Newberry Park.*

Mooney, C. Z., and Krause, G. (1998). Of silicon and political science: Computationally intensive techniques of statistical estimation and inference. *Br. J. Political Sci.* **27**, 83–110.

Moreau, J. V., and Jain, A. K. (1987). The bootstrap approach to clustering. In *Pattern Recognition: Theory and Applications* (P. A. Devijver and J. Kittler, editors), pp. 63–71. Springer-Verlag, Berlin.

Morey, M. J., and Schenck, L. M. (1984). Small sample behavior of bootstrapped and jackknifed regression estimators. In *Proceedings, ASA Business Economics and Statistics Section*, pp. 437–442.

Morgenthaler, S., and Tukey, J. W. (editors). *Configural Polysampling: A Route to Practical Robustness*. Wiley, New York.

Morris, M. D., and Ebey, S. F. (1984). An interesting property of the sample mean under a first-order autoregressive model. *Amer. Statist.* **38**, 127–129.

Morton, S. C. (1990). Bootstrap confidence intervals in complex situation: A sequential paired clinical trial. *Commun. Statist. Simul. Comput.* **19**, 181–195.

Mosbach, O. (1988). Bootstrap-verfahren in allgemeinen linearen modellen. Universitat Bremen, Fachbereich 03: Diplomarbeit.

Mosbach, O. (1992). One-step bootstrapping in generalized linear models. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 143–147. Springer-Verlag, Berlin.

Moosman, D. (1995). Resampling techniques in the analysis of non-binomial ROC data. *Med. Dec. Making* **15**, 358–366.

Moulton, L. H., and Zeger, S. L. (1991). Bootstrapping generalized linear models. *Comput. Statist. Data Anal.* **11**, 53–63.

Moulton, L. H., and Zeger, S. L. (1989). Analyzing repeated measures on generalized linear models via the bootstrap. *Biometrics* **45**, 381–394.

Mueller, L. D., and Altenberg, L. (1985). Statistical inference on measures of niche overlap. *Ecology* **66**, 1204–1210.

Mueller, L. D., and Wang, J. L. (1990). Bootstrap confidence intervals for effective doses in the probit model for dose–response data. *Biom. J.* **32**, 529–544.

Mueller, P. (1986). On selecting the set of regressors. In *Classification as a Tool of Research* (W. Gaul and M. Schader, editors), pp. 331–338. North-Holland, Amsterdam.

Murthy, V. K. (1974). *The General Point Process: Applications to Structural Fatigue, Bioscience and Medical Research*. Addison-Wesley, Boston.*

Mykland, P. (1992). Asymptotic expansions and bootstrapping distribution for dependent variables: A martingale approach. *Ann. Statist.* **20**, 623–654.

Myoungshic, J. (1986). Bootstrap method for *K*-spatial medians. *J. Korean Statist.* **15**, 1–8.

Nagao, H. (1985). On the limiting distribution of the jackknife statistics for eigenvalues of a sample covariance matrix. *Commun. Statist. Theory Methods* **14**, 1547–1567.

Nagao, H. (1988). On the jackknife statistics for eigenvalues and eigenvectors of a correlation matrix. *Ann. Inst. Statist. Math.* **40**, 477–489.

Nagao, H., and Srivastava, M.S. (1992). On the distributions of some test criteria for a covariance matrx under local alternatives and bootstrap approximations. *J. Multivar. Anal.* **43**, 331–350.

Navidi, W. (1989). Edgeworth expansions for bootstrapping in regression models. *Ann. Statist.* **17**, 1472–1478.

Navidi, W. (1995). Bootstrapping a method of phylogenetic inference. *J. Statist. Plann. Inf.* **43**, 169–184.

Ndlovu, P. (1993). Classical and bootstrap estimates of heritability of milk yield in Zimbabwean Holstein cows. *J. Dairy Sci.* **76**, 2013–2024.

Nelder, J. A. (1996). Statistical computing. In *Advances in Biometry* (P. Armitage, and H. A. David, editors), pp. 201–212. Wiley, New York.

Nelson, L. S. (1994). Book Review of *Permutation Tests* by P. I. Good, *J. Qual. Technol.* **26**, 325.

Nelson, R. D. (1992). Applications of stochastic dominance using truncation, bootstrapping and kernels. *Am. Statist. Assoc. Proc. Bus. Econ. Statist. Sect.* 88–93.

Nelson, W. (1990). *Accelerated Testing: Statistical Models, and Data Analyses*. Wiley, New York.*

Nemec, A. F. L., and Brinkhurst, R. O. (1988). Using the bootstrap to assess statistical significance in the cluster analysis of species abundance data. *Can. J. Fish. Aquat. Sci.* **45**, 971–975.

Newton, M. A., and Geyer, G. J. (1994). Bootstrap recycling: A Monte Carlo alternative to the nested bootstrap. *J. Am. Statist. Assoc.* **89**, 905–912.*

Newton, M. A., and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *J. R. Statist. Soc. B* **56**, 3–48.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.*

Nigam, A. K., and J. N. K. Rao (1996). On balanced bootstrap for stratified multistage samples. *Statist. Sin.* **6**, 199–214.*

Nirel, R. (1994). Bootstrap confidence intervalsfor the estimation of seeding effect in an operational period. *Statist. Environ.* **2**, 109–124.

Nishizawa, O., and Noro, H. (1995). Bootstrap statistics for velocity tomography: Application of a new information criterion. *Geophys. Prospect.* **43**, 157–176.

Nivelle, F., Rouy, V., and Vergnaud, P. (1993). Optimal design of neural network using resampling methods. In *Proceedings of the 6th International Conference, Neural Networks and Their Industrial Cognitive Applications*, pp. 95–106.

Nokkert, J. H. (1985). Comparison of confidence regions for a linear functional relationship based on moment-jackknife-and bootstrap estimators of an asymptotic covariance matrix. *Statist. Dec. Sci.* **I.2**, 207–212.

Nordgaard, A. (1990). On the resampling of stochastic processes using the bootstrap approach. Liu-Tek-Lic-**1990:23**, Linkoping, Sweden.

Nordgaard, A. (1992). Resampling stochastic processes using a bootstrap approach. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), pp. 376, 181–183. Springer-Verlag, Berlin.

Noreen, E. (1989). *Computer-Intensive Methods for Testing Hypotheses*. Wiley, New York.*

Nychka, D. (1991). Choosing a range for the amount of smoothing in nonparametric regression. *J. Am. Statist. Assoc.* **86**, 653–664.

Oakley, E. H. N. (1996). Genetic programming , the reflection of chaos and the bootstrap: Toward a useful test for chaos. In *Genetic Programming, Proceedings of the 1st Annual Conference 1996* (J. R. Koza, D. E. Goldberg, D. B. Dogel, and R. Riolo, editors), pp. 175–181.

Obgonmwan, S.-M. (1985). Accelerated resampling codes with applications to likelihood. Ph.D. Thesis, Department of Mathematics, Imperial College, London.*

Obgonmwan, S.-M., and Wynn, H. P. (1986). Accelerated resampling codes with low discrepancy. Department of Statistics and Actuarial Science, City University, London.

Obgonmwan, S.-M., and Wynn, H. P. (1988). Resampling generated likelihoods. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger editors), pp. 133–147. Springer-Verlag, New York.

Oden, N. L. (1991). Allocation of effort in Monte Carlo simulation for power of permutation tests. *J. Am. Statist. Assoc.* **86**, 1007–1012.

Oldford, R. W. (1985). Bootstrapping by Monte Carlo versus approximating the estimator and bootstrapping exactly. *Commun. Statist. Simul. Comput.* **14**, 395–424.

Olshen, R. A., Biden, E. N., Wyatt, M. P., and Sutherland, D. H. (1989). Gait analysis and the bootstrap. *Ann. Statist.* **17**, 1419–1440.*

Ooms, M., and Franses, P. H. (1997). On periodic correlations between estimateeed seasonal and nonseasonal components in German and U.S. unemployment. *J. Bus. Econ. Statist.* **15**, 470–481.

O'Quigley, J., and Pessione, F. (1991). The problem of a covariate-time qualitative interaction survival study. *Biometrics* **47**, 101–115.

O'Sullivan, F. (1988). Parameter estimation in parabolic and hyperbolic equations. Technical Report 127, Department of Statistics, University of Washington, Seattle.

Overton, W. S., and Stehman, S. V. (1994). Improvements of performance of variable probability sampling strategies through application of the population space and facsimile population bootstrap. Oregon State University, Department of Statistics Technical Report.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.

Owen, A. B. (1992). A central limit theorem for Latin hypercube sampling. *J. R. Statist. Soc. B* **54**, 541–551.*

Paas, G. (1994). Assessing predictive accuracy by the bootstrap algorithm. In *Proceedings of the International Conference on Artificial Neural Networks, ICANN '94* (M. Marinaro, and P. G. Morasso, editors), Vol. 2, pp. 823–826.

Padgett, W. J., and Thombs, L. A. (1986). Smooth nonparametric quantile estimation under censoring: Simulations and bootstrap methods. *Commun. Statist. Simul. Comput.* **15**, 1003–1025.

Padmanabhan, A. R., Chinchilli, V. M., and Babu, G. J. (1997). Robust analysis of within-unit variances in repeated measurement experiments. *Biometrics* **53**, 1520–1526.

Paez, T. L. , and Hunter, N. F. (1998). Fundamental concepts of the bootstrap for statistical analysis of mechanical systems. *Exp. Techniques* **22**, 35–38.

Page, J. T. (1985). Error-rate estimation in discriminant analysis. *Technometrics* **27**, 189–198.*

Pallini, A., Carletti, M., and Pesarin, F. (1994). *Proc. Ital. Statist. Soc.* **2**, 441–448.

Pallini, A., and Pesarin, F. (1994). Calibration resampling for the conditional bootstrap. *Proc. Ital. Statist. Soc.* **2**, 473–480.

Papadopoulos, A. S., and Tiwari, R. C. (1989). Bayesian Bootstrap lower confidence interval estimation of the reliability and failure rate. *J. Statist. Comput. Simul.* **32**, 185–192.

Paparoditis, E. (1992). *Bootstrapping* some statistics useful in identifying ARMA models. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler editors), Vol. 376, pp. 115–119. Springer-Verlag, Berlin.

Paparoditis, E. (1995). A frequency domain bootstrap-based method for checking the fit of a transfer function model. *J. Am. Statist. Assoc.* **91**, 1535–1550.

Pari, R., and Chatterjee, S. (1986). Using L_2 estimation for L_1 estimators: An application of the single-index model. *Dec. Sci.* **17**, 414–423.

Parmanto, B., Munro, P. W., and Doyle, H. R. (1996a). Improving committee diagnosis with resampling techniques. In *Proceedings of the 1995 Conference: Advances in Neural Information Processing Systems* (D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors), Vol. 8, pp. 882–888.

Parmanto, B., Munro, P. W., and Doyle, H. R. (1996b). Reducing variance of committee prediction with resampling techniques. *Connect. Sci.* **8**, 405–425.

Parr, W. C. (1983). A note on the jackknife, the bootstrap and the delta method estimators of bias and variance. *Biometrika* **70**, 719–722.*

Parr. W. C. (1985a). Jackknifing differentiable statistical functionals. *J. Roy. Statist. Soc. B* **47**, 56–66.

Parr. W. C. (1985b). The bootstrap: Some large sample theory and connections with robustness. *Statist. Probab. Lett.* **3**, 97–100.

Parzen, E. (1982). Data modeling using quantile and density-quantile functions. In *Some Recent Advances in Statistics* (J. T. de Oliveira and B. Epstein, editors), pp. 23–52. Academic Press, London.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

Pavia, E. G., and O'Brien, J. J. (1986). Weibull statistics of wind speed over the ocean. *J. Clim. A. Mtr.* **25**, 1324–1332.

Peck, R., Fisher, L., and Van Ness, J. (1989). Approximate confidence intervals for the number of clusters. *J. Am. Statist. Assoc.* **84**, 184–191.

Pederson, S. P., and Johnson, M. E. (1990). Estimating model discrepancy. *Technometrics* **32**, 305–314.

Peladeau, N., and Lacouture, Y. (1993). SIMSTAT: Bootstrap computer simulation and statistical program for IBM personal computers. *Behav. Res. Methods, Instrum. Comput.* **25**, 410–413.

Peters, S. C., and Freedman, D. A. (1984a). Bootstrapping an econometric model: Some empirical results. *J. Bus. Econ. Statist.* **2**, 150–158.

Peters, S. C., and Freedman, D. A. (1984b). Some notes on the bootstrap in regression problems. *J. Bus. Econ. Statist.* **2**, 401–409.*

Peters, S. C., and Freedman, D. A. (1985). Using the bootstrap to evaluate forecast equations. *J. Forecast.* **4**, 251–262.*

Peters, S. C., and Freedman, D. A. (1987). Balm for bootstrap confidence intervals. *J. Am. Statist. Assoc.* **82**, 186–187.

Peterson, A. V. (1983). Kaplan–Meier estimator. In *Encyclopedia of Statistical Sciences*, Vol. 4. Wiley, New York.

Pettit, A. N. (1987). Estimates for a regression parameter using ranks. *J. R. Statist. Soc. B* **49**, 58–67.

Pewsey, A. (1994). Book Review of *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors). *Statistician* **43**, 215–216.

Pham, T. D., and Nguyen, H. T. (1993). Bootstrapping the change-point of a hazard rate. *Ann. Inst. Statist. Math.* **45**, 331–340.

Picard, R. R., and Berk, K. N. (1990). Data splitting. *Am. Statist.* **44**, 140–147.

Pictet, O. V., Dacorogna, M. M., and Muller, U. A. (1998). Hill bootstrap and jackknife estimators for heavy tails. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors), pp. 283–310. Birkhauser, Boston.*

Pigeot, I. (1991a). A jackknife estimator of a common odds ratio. *Biometrics* **47**, 373–381.

Pigeot, I. (1991b). A simulation study of estimators of a common odds in several 2×2 tables. *J. Statist. Comput. Simul.* **38**, 65–82.

Pigeot, I. (1992). Jackknifing estimators of a common odds ratio from several 2×2 tables. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems*, (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 204–212. Springer-Verlag, Berlin.

Pigeot, I. (1994). Special resampling techniques in categorical Sara analysis. In *the 25th Conference on Statistical Computing: Computational Statistics* (P. Dirschedl, and R. Ostermann, editors), pp. 159–176. Physica-Verlag, Heidelberg.

Pinheiro, J. C., and DeMets, D. L. (1997). Estimating and reducing bias in group sequential designs with Gaussian independent increment structure. *Biometrika* **84**, 831–845.

Pitt, D. G., and Kreutzweiser, D. P. (1998). Applications of computer-intensive statistical methods to environmental research. *Ecotoxicol. Environ. Saf.* **39**, 78–97.

Platt, C. A. (1982). Bootstrap stepwise regression. In *Proceedings, Business Economics and Statistics Section of the American Statistical Association*, pp. 586–589.

Plotnick, R. E. (1989). Application of bootstrap methods to reduced major axis line fitting. *Syst. Zool.* **38**, 144–153.

Politis, D. N. (1998). Computer-intensive methods in statistical analysis. *IEEE Signal Process. Mag.* **15**, 39–55.*

Politis, D. N., and Romano, J. P. (1990). A nonparametric resampling procedure for multivariate confidence regions in time series analysis. In *Proceedings of INTERFACE' 90, 22nd Symposium on the Interface of Computer Science and Statistics* (C. Page and R. LePage, editors). 98–103. Springer-Verlag, New York.

Politis, D. N., and Romano, J. P. (1992a). A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 263–270. Wiley, New York.*

Politis, D. N., and Romano, J. P. (1992b). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.* **20**, 1985–2007.

Politis, D. N., and Romano, J. P. (1993a). Estimating the distribution of a studentized statistic by subsampling. *Bull. Intern. Statist. Inst. 49th Session* **2**, 315–316.*

Politis, D. N., and Romano, J. P. (1993b). Nonparametric resampling for homogeneous strong mixing random fields. *J. Mult. Anal.* **47**, 301–328.

Politis, D. N., and Romano, J. P. (1994a). The stationary bootstrap. *J. Am. Statist. Assoc.* **89**, 1303–1313.*

Politis, D. N., and Romano, J. P. (1994b). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031–2050.*

Politis, D. N., and Romano, J. P. (1994c). Limit theorems for weakly dependent Hilbert space valued random variables with applications to the stationary bootstrap. *Statist. Sin.* **4**, 461–476.

Politis, D. N., Romano, J. P., and Lai, T. L. (1992). Bootstrap confidence bands for spectra and cross-spectra. *IEEE Trans. Signal Process.* **40**, 1206–1215.*

Politis, D. N., Paparoditis, E., and Romano, J. P. (1998). Large sample inference for irregularly spaced dependent observations based on subsampling. *Sankhya A* **60**, 274–292.

Politis, D. N., Romano, J. P., and Wolf (1999).

Pollack, S., Bruce, P., Borenstein, M., and Lieberman, J. (1994). The resampling method of statistical analysis. *Psychopharmacol. Bull.* **30**, 227–234.

Pollack, S., Simon, J., Bruce, P., Borenstein, M., and Lieberman, J. (1994). Using teaching, and evaluating the resampling method of statistical analysis. *Psychopharmacol. Bull.* **30**, 120.*

Pons, O., and de Turckheim, E. (1991a). von Mises methods, bootstrap and Hadamard differentiability for nonparametric general models. *Statistics* **22**, 205–214.

Pons, O., and de Turckheim, E. (1991b). Tests of independence for bivariate censored data based on the empirical joint hazard function. *Scand. J. Statist.* **18**, 21–37.

Portnoy, S. (1984). Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles. In *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Hardle, and R. D. Martin, editors), pp. 231–246. Springer-Verlag, New York.

Prada-Sanchez, J. M., and Cotos-Yanez, T. (1997). A simulation study of iterated and non-iterated bootstrap methods for bias reduction and confidence interval estimation. *Commun. Statist. Simul. Comput.* **26**, 927–946.

Praestgaard, J., and Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21**, 2053–2086.

Presnell, B., and Booth, J. G. (1994). Resampling methods for sample surveys. Technical Report 470, Department of Statistics, University of Florida, Gainsville.*

Presnell, B., Morrison, S. P., and Littell, R. C. (1998). Projected multivariate linear models for directional data. *J. Am. Statist. Assoc.* **93**, 1068–1077.

Press, S. J. (1989). *Bayesian Statistics: Principles, Models, and Applications*. Wiley, New York.*

Price, B., and Price, K. (1992). Sampling variability of capability indices. Technical Report, Wayne State University, Detroit.*

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.*

Proenca, I. (1990). Metodo bootstrap—Uma aplicacao na estimacao e previsao em modelos dinamicos. Unpublished M.Sc. dissertation, ISEG, Lisboa.

Pugh, G. A. (1995). Resampled confidence bounds on effects diagrams for signal-to-noise. *Comput. Ind. Eng.* **29**, 11–13.

Qin, J., and Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika* **84**, 609–618.

Quan, H., and Tsai, W.-Y. (1992). Jackknife for the proportional hazards model. *J. Statist. Comput. Simul.* **43**, 163–176.

Quenneville, B. (1986). Bootstrap procedures for testing linear hypotheses without normality. *Statistics* **17**, 533–538.

Quenouille, M. H. (1949). Approximate tests of correlation in time series. *J. Roy. Statist. Soc. B* **11**, 18–84.*

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353–360.

Racine, J. (1997). Consistent significance testing for nonparametric regression. *J. Bus. Econ. Statist.* **15**, 369–378.

Raftery, A. E. (1995). Hypothesis testing and model selection. In *Markov Chain Monte Carlo Practice* (W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors), pp. 163–187. Chapman & Hall, London.

Rajarshi, M. B. (1990). Bootstrap in Markov sequences based on estimate of transition density. *Ann. Inst. Statist. Math.* **42**, 253–268.

Ramos, E. (1988). Resampling methods for time series. Ph.D. thesis, Department of Statistics, Harvard University.

Rao, C. R., Pathak, P. K., and Koltchinskii, V. I. (1997). Bootstrap by sequential resampling. *J. Statist. Plann. Inf.* **64**, 257–281.

Rao, C. R. (1982). Diversity, its measurement, decomposition, apportionment and analysis. *Sankhya A* **44**, 1–21.

Rao, C. R., and Zhao, L. (1992). Approximation to the distribution of *M*-estimates in linear models by randomly weighted bootstrap. *Sankhya A* **54**, 323–331.

Rao, J. N. K., Kovar, J. G., and Mantel, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika* **77**, 365–375.

Rao, J. N. K., and Wu, C. F. J. (1985). Inference from stratified samples: Second-order analysis of three methods for nonlinear statistics. *J. Am. Statist. Assoc.* **80**, 620–630.

Rao, J. N. K., and Wu, C. F. J. (1987). Methods for standard errors and confidence intervals from sample survey data: Some recent work. *Bull. Int. Statist. Inst.* **52**, 5–21.

Rao, J. N. K., and Wu, C. F. J. (1988). Resampling inference with complex survey data. *J. Am. Statist. Assoc.* **83**, 231–241.*

Rao, J. N. K., Wu, C. F. J., and Yuen, K. (1992). Some recent work on resampling methods for complex surveys. *Surv. Methodol.* **18**, 209–217.

Rasmussen, J. L. (1987a). Estimating correlation coefficients: bootstrap and parametric approaches. *Psych. Bull.* **101**, 136–139.

Rasmussen, J. L. (1987b). Parametric and bootstrap approaches to repeated measures designs. *Behav. R. Me. In.* **19**, 357–360.

Raudys, S. (1988). On the accuracy of a bootstrap estimate of classification error. In *Proceedings on the 9th International Conference on Pattern Recognition*, pp. 1230–1232.

Rayner, R. K. (1990a). Bootstrapping *p* values and power in the first-order autoregression: A Monte Carlo investigation. *J. Bus. Econ. Statist.* **8**, 251–263.*

Rayner, R. K. (1990b). Bootstrap tests for generalized least squares regression models. *Econom. Lett.* **34**, 261–265.*

Rayner, R. K., and Dielman, T. E. (1990). Use of the bootstrap in tests for serial correlation when regressors include lagged dependent variables. Unpublished Technical Report.*

Red-Horse, J. R., and Paez, T. L. (1998). Assessment of probability models using the bootstrap sampling method. *39th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference and Exhibit, Collections of Technical Papers, Part 2, AIAA*, 1086–1091.

Reiczigel, J. (1996). Bootstrap tests in correspondence analysis. *Appl. Stoch. Models Data Anal.* **12**, 107–117.

Reid, N. (1981). Estimating the median survival time. *Biometrika* **68**, 601–608.*

Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* **3**, 213–238.*

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics*. Springer-Verlag, New York.*

Reiss, R.-D., and Thomas, M. (1997). *Statistical Analysis on Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhauser Verlag, Basel.*

Reneau, D. M., and Samaniego, F. J. (1990). Estimating the survival curve when new is better than used of a specified age. *J. Am. Statist. Assoc.* **85**, 123–131.

Resampling Stats., Inc. (1997). *Resampling Stats User's Guide* (P. Bruce, J. Simon, and T. Oswald, authors).*

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York.*

Resnick, S. I. (1997). Heavy tail modeling and teletraffic data (with discussion). *Ann. Statist.* **25**, 1805–1869.

Rey, W. J. J. (1983). *Introduction to Robust and Quasi-Robust Statistical Methods*. Springer-Verlag, Berlin.*

Rice, R. E., and Moore, A. H. (1983). A Monte-Carlo technique for estimating lower confidence limits on system reliability using pass-fail data. *IEEE Trans. Rel.* **R-32**, 366–369.

Rieder, H. (editor) (1996). *Robust Statistics, Data Analysis, and Computer Intensive Methods*. Lecture Notes in Statistics, Vol. 109. Springer-Verlag, Heidelberg.

Riemer, S., and Bunke, O. (1983). A note on bootstrap and other empirical procedures for testing linear hypotheses without normality. *Statistics* **14**. 517–526.

Ringrose, T. J. (1994). Bootstrap confidence regions for canonical variate analysis. In *Proceedings in Computational Statistics, 11th Symposium*, pp. 343–348.

Ripley, B. D. (1981). *Spatial Statistics*. Wiley, New York.*

Ripley, B. D. (1987). *Stochastic Simulation*. Wiley, New York.*

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.*

Ripley, B. D. (1992). Applications of Monte Carlo methods in spatial and image analysis. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems*, (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 47–53. Springer-Verlag, Berlin.

Rissanen, J. (1986). Stochastic complexity and modeling. *Ann. Statist.* **14**, 1080–1100.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*. World Scientific Publishing Co., Singapore.*

Roberts, F. S. (1984). *Applied Combinatorics*. Prentice-Hall, Englewood Cliffs.*

Robeson, S. M. (1995). Resampling of network-induced variability in estimates of terrestrial air temperature change. *Climate Change* **29**, 213–229.*

Robinson, J. A. (1983). Bootstrap confidence intervals in location-scale model with progressive censoring. *Technometrics* **25**, 179–187.

Robinson, J. A. (1986). Bootstrap and randomization confidence intervals. In *Pacific Statistical Congress: Proceedings of the Congress* (I. S. Francis, B. F. J. Manly, and F. C. Lam, editors), pp. 49–50, North-Holland, Amsterdam.

Robinson, J. A. (1987). Nonparametric confidence intervals in regression, the bootstrap and randomization methods. In *New Perspectives in Theoretical and Applied Statistics* (M. Puri, J. P. Vilaplana, and W. Wertz, editors), pp. 243–256. Wiley, New York.

Robinson, J. A. (1994). Book review of *An Introduction to the Bootstrap* (by B. Efron and R. J. Tibshirani). *Austral. J. Statist.* **36**, 380–382.

Robinson, J., Feuerverger, A., and Jing, B.-Y. (1994) On the bootstrap saddlepoint approximations. *Biometrika* **81**, 211–215.

Rocke, D. M. (1989). Bootstrapping Bartlett's adjustment in seemingly unrelated regression. *J. Am. Statist. Assoc.* **84**, 598–601.

Rocke, D. M. (1993). Almost-exact parametric bootstrap calibration via the saddlepoint approximation. *Comput. Statist. Data Anal.* **15**, 179–198.

Rocke, D. M., and Downs, G. W. (1981). Estimating the variance of robust estimators of location: Influence curve, jackknife and bootstrap. *Commun. Statist. Simul. Comput.* **10**, 221–248.

Rodriguez-Campos, M. C., and Cao-Abad, R. (1993). Nonparametric bootstrap confidence intervals for discrete regression functions. *J. Econometrics* **58**, 207–222.

Romano, J. P. (1988a). A bootstrap revival of some nonparametric distance tests. *J. Am. Statist. Assoc.* **83**, 698–708.*

Romano, J. P. (1988b). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16**, 629–647.

Romano, J. P. (1988c). Bootstrapping the mode. *Ann. Inst. Statist. Math.* **40**, 565–586.*

Romano, J. P. (1989a). Do bootstrap confidence procedures behave well uniformly in P? *Can. J. Statist.* **17**, 75–80.

Romano, J. P. (1989b). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17**, 141–159.*

Romano, J. P. (1990). On the behavior of randomization tests of some nonparametric hypotheses. *J. Am. Statist. Assoc.* **85**, 686–692.

Romano, J. P., and Siegel, A. F. (1986). *Counterexamples in Probability and Statistics*. Wadsworth, Monterey.

Romano, J. P., and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Am. Statist. Assoc.* **91**, 590–600.

Rosa, R., and Mignani, S. (1994). Moving block bootstrap for dependent data: An application to Ising models (Italian). *Proc. Ital. Statist. Soc.* **2**, 465–472.

Ross, S. M. (1990). *A Course in Simulation*. Macmillan, New York.

Rothe, G. (1986a). Some remarks on bootstrap techniques for constructing confidence intervals. *Statist. Hefte* **27**, 165–172.

Rothe, G. (1986b). Bootstrap in generalisierten linearen modellen. *ZUMA-Arbeitsbericht* **86–11**, Mannheim.

Rothe, G., and Armingeer, G. (1992). Bootstrap for mean and covariance structure models. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems*, (K.-H. Jockel, G. Rothe and W. Sendler, editors), Vol. 376, pp. 149–155. Springer-Verlag, Berlin.

Rothery, P. (1985). Estimation of age-specific survival in hen harriers (*Circus cyaneus*) in Orkney. In *Statistics in Ornithology* (B. J. T. Morgan, and P. M. North, editors), pp. 341–354. Springer-Verlag, Berlin.*

Rousseeuw, P. J. (1984). Least median of squares regression. *J. Am. Statist. Assoc.* **79**, 872–880.*

Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.*

Rousseeuw, P. J., and Yohai, V. (1984). Robust regression by means of *s*-estimators. In *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Hardle, and R. D. Martin, editors), pp. 256–272. Springer-Verlag, New York.

Roy, T. (1994). Bootstrap accuracy for non-linear regression models. *J. Chemometrics* **8**, 37–44.*

Roystone, P., and Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling. *Appl. Statist.* **43**, 429–467.*

Rubin, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9**, 130–134.*

Rubin, D. B. (1983). A case study of the robustness of Bayesian methods of inference: Estimating the total in a finite population using transformations to normality. In *Scientific Inference, Data Analysis, and Robustness* (G. E. P. Box, T. Leonard, and C.-F. Wu. editors), pp. 213–244. Academic Press, London.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.*

Rubin, D. B. (1996). Multiple imputation after 18+ years. *J. Am. Statist. Assoc.* **91**, 473–489.

Rubin, D. B., and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J. Am. Statist. Assoc.* **81**, 366–374.*

Rubin, D. B., and Schenker, N. (1991). Multiple imputation in health-care data bases: An overview and some applications. *Statist. Med.* **10**, 585–598.*

Rubin, D. B., and Schenker, N. (1998). Imputation. In *Encyclopedia of Statistical Sciences*, Update Volume 2 (S. Kotz, C. B. Read, and D. L. Banks, editors), pp. 336–342. Wiley, New York.*

Runkle, D. E. (1987). Vector autoregressions and reality (with discussion). *J. Bus. Econ. Statist.* **5**, 437–453.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *J. Off. Statist.* **1**, 381–397.

Ryan, T. P. (1989). *Statistical Methods for Quality Improvement*. Wiley, New York.*

Sager, T. W. (1986). Dimensionality reduction in density estimation. In *Statistical Image Processing and Graphics* (E. J. Wegman, and D. J. DePriest, editors), pp. 307–319. Marcel Dekker, New York.

Sain, S. R., Baggeily, K. A., and Scott, D. W. (1994). Cross-validation of multivariate densities. *J. Am. Statist. Assoc.* **89**, 807–817.

Samawi, H. M. (1994). Power estimation for two-sample tests using importance and antithetic resampling. Ph.D. thesis. Department of Actuarial Science, University of Iowa.

Sanderson, M. J. (1989). Confidence limits on phylogenies: The bootstrap revisited. *Cladistics* **5**, 113–129.*

Sanderson, M. J. (1995). Objections to bootstrapping phylogenies: A critique. *Syst. Biol.* **44**, 299–320.*

Sauerbrei, W., and Schumacher, M. (1992). A bootstrap resampling procedure for model building: Application to the Cox regression model. *Statist. Med.* **11**, 2093–2109.*

Sauerbrei, W. (1998). Bootstrapping in survival analysis. In *Encyclopedia of Biostatistics* (P. Armitage and T. Colton, editors), Vol. 1, pp. 433–436. Wiley, New York.*

Sauermann, W. (1986). Bootstrap—Verfahren in log-linearen Modellen. Dissertation. Universitat Heidelberg.

Sauermann, W. (1989). Bootstrapping the maximum likelihood estimator in high-dimensional log-linear models. *Ann. Statist.* **17**, 1198–1216.

Schader, R. M., and McKean, J. W. (1987). Small sample properties of least absolute errors. Analysis of variance. In *Statistical Data Analysis Based on the L_1 Norm and Related Methods* (Y. Dodge, editor), pp. 307–321. North-Holland, Amsterdam.

Schafer, H. (1992). An application of the bootstrap in clinical chemistry. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 213–217. Springer-Verlag, Berlin.*

Schall, R., and Luus, H. G. (1993). On population and individual bioequivalence. *Statist. Med.* **12**, 1109–1124.

Schemper, M. (1987a). Nonparametric estimation of variance, skewness and kurtosis of the distribution of a statistic by jackknife and bootstrap techniques. *Statist. Neerl.* **41**, 59–64.

Schemper, M. (1987b). On bootstrap confidence limits for possibly skew distributed statistics. *Commun. Statist. Theory Methods* **16**, 1585–1590.

Schemper, M. (1987c). One- and two-sample tests of Kendall's τ. *Biometric J.* **29**, 1003–1009.

Schenker, N. (1985). Qualms about bootstrap confidence intervals. *J. Am. Statist. Assoc.* **80**, 360–361.*

Schervish, M. J. (1995). *Theory of Statistics*. Springer-Verlag, New York.*

Schork, N. (1992). Bootstrapping likelihood ratios in quantitative genetics. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard editors), pp. 389–396. Wiley, New York.*

Schluchter, M. D., and Forsythe, A. B. (1986). A caveat on the use of a revised bootstrap algorithm. *Psychometrika* **51**, 603–605.

Schucany, W. R. (1988). Sample reuse. In *Encyclopedia of Statistical Sciences*, Vol. 8, pp. 235–238. Wiley, New York.

Schucany, W. R., and Bankson, D. M. (1989). Small sample variance estimators for *U*-statistics. *Aust. J. Statist.* **31**, 417–426.

Schucany, W. R. and Sheather, S. J. (1989). Jackknifing *R*-estimators. *Biometrika* **76**, 393–398.

Schucany, W. R., and Wang, S. (1991). One-step bootstrapping for smooth iterative procedures. *J. R. Statist. Soc. B* **53**, 587–596.

Schumacher, M., Hollander, N., and Sauerbrei,W. (1997). Resampling and cross-validation techniques. A tool to reduce bias caused by model building? *Statist. Med.* **16**, 2813–2828.

Schuster, E. F. (1987). Identifying the closest symmetric distribution or density function. *Ann. Statist.* **15**, 865–874.

Schuster, E. F., and Barker, R. C. (1989). Using the bootstrap in testing symmetry versus asymmetry. *Commun. Statist. Simul. Comput.* **16**, 69–84.

Scott, D. W. (1992). *Multivariate Density Estimation, Theory, Practice, and Visualization*, Wiley, New York.

Seber, G. A. F. (1984). *Multivariate Observations*. Wiley, New York.*

Seki, T., and Yokoyama, S. (1996). Robust parameter-estimation using the bootstrap method for the 2-parameter Weibull distribution. *IEEE Trans. Reliab.* **45**, 34–41.

Sen, A., and Srivastiva, M. S. (1990). *Regression Analysis: Theory, Methods, and Applications*. Springer-Verlag, New York.*

Sen, P. K. (1988a). Functional approaches in resampling plans: A review of some recent developments. *Sankhya A* **50**, 394–435.

Sen, P. K. (1988b). Functional jackknifing: Rationality and general asymptotics. *Ann. Statist.* **16**, 450–469.*

Seppala, T., Moskowitz, H., Plante, R., and Tang, J. (1995). Statistical process control via the subgroup bootstrap. *J. Qual. Tech.* **27**, 139–153.*

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.*

Sezgin, N., and Bar-Ness, Y. (1996). Adaptive soft limiter bootstrap separator for one-shot asynchronous CDMA channel with singular partial cross-correlation matrix. In *IEEE International s, Convergung Conference on Communications, Converging Technologies for Tomorrow's Applications. ICC '96*, **1**, 73–77.

Shao, J. (1987a). Bootstrap variance estimations. Ph.D. dissertation, Department of Statistics, University of Wisconsin, Madison.

Shao, J. (1987b). Sampling and resampling: An efficient approximation to jackknife variance estimators in linear models. *Ch. J. Appl. Probab. Statist.* **3**, 368–379.

Shao, J. (1988a). On resampling methods for variance and bias estimation in linear models. *Ann. Statist.* **16**, 986–1008.*

Shao, J. (1988b). Bootstrap variance and bias estimation in linear models. *Can. J. Statist.* **16**, 371–382.*

Shao, J. (1989). Bootstrapping for generalized *L*-statistics. *Commun. Statist. Theory Methods* **18**, 2005–2016.

Shao, J. (1990a). Influence function and variance estimation. *Ch. J. Appl. Probab. Statist.* **6**, 309–315.

Shao, J. (1990b). Bootstrap estimation of the asymptotic variance of statistical functionals. *Ann. Inst. Statist. Math.* **42**, 737–752.

Shao, J. (1991). Consistency of jackknife variance estimators. *Statistics* **22**, 49–57.

Shao, J. (1992). Bootstrap variance estimators with truncation. *Statist. Probab. Lett.* **15**, 95–101.

Shao, J. (1993). Linear model selection by cross-validation. *J. Am. Statist. Assoc.* **88**, 486–494.

Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proc. Am. Math. Soc.* **122**, 1251–1262.

Shao, J. (1996). Bootstrap model selection. *J. Am. Statist. Assoc.* **91**, 655–665.*

Shao, J., and Sitter, R. R. (1996). Bootstrap for imputed survey data. *J. Am. Statist. Assoc.* **91**, 1278–1288.*

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.*

Shao, J., and Wu, C. F. J. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist.* **15**, 1563–1579.*

Shao, J., and Wu, C. F. J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17**, 1176–1197.

Shao, J., and Wu, C. F. J. (1992). Asymptotic properties of the balanced repeated replication method for sample quantiles. *Ann. Statist.* **20**, 1571–1593.

Shao, Q., and Yu. H. (1993). Bootstrapping the sample means for stationary mixing sequences. *Stocn. Proc. and Their Appl.* **48**, 175–190.*

Shaw, F. H., and Geyer, C. J. (1997). Estimation and testing in constrained covariance component models. *Biometrika* **84**, 95–102.

Sheather, S. J. (1987). Assessing the accuracy of the sample median: Estimated standard errors versus interpolated confidence intervals. In *Statistical Data Analysis Based on the L_1 Norm and Related Methods* (Y. Dodge, editor), pp. 203–215. North-Holland, Amsterdam.

Sheather, S. J., and Marron, J. S. (1990). Kernel quantile estimators. *J. Am. Statist. Assoc.* **85**, 410–416.

Shen, C. F., and Iglewicz, B. (1994). Robust and bootstrap testing procedures for bio-equivalence. *J. Biopharmacol. Stat.* **4**, 65–90.*

Sherman, M. (1997). Subseries methods in regression. *J. Am. Statist. Assoc.* **92**, 1041–1048.

Sherman, M., and Carlstein, E. (1996). Replicate histograms. *J. Am. Statist. Assoc.* **91**, 566–576.

Sherman, M. and Carlstein, E. (1997). Omhibus Confidence Internals. Technical Report, Department of Statistics. Texas A & M University.*

Sherman, M., and le Cessie, S. (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. *Commun. Statist. Simul. Comput.* **26**, 901–925.

Shi, X. (1984). The approximate independence of jackknife pseudo-values and the bootstrap methods. *J. Wuhan Inst. Hydra. Elect. Eng.* **2**, 83–90.

Shi, X. (1986a). A note on bootstrapping *U*-statistics. *Ch. J. Appl. Probab. Statist.* **2**, 144–148.

Shi, X. (1986b). Bootstrap estimate for *m*-dependent sample means. *Kexue Tongbao (Chin. Bull. Sci.)* **31**, 404–407.

Shi, X. (1987). Some asymptotic properties of bootstrapping *U*-statistics. *J. Syst. Sci. Math. Sci.* **7**, 23–26.

Shi, X. (1991). Some asymptotic results for jackknifing the sample quantile. *Ann. Statist.* **19**, 496–503.

Shi, X., and Liu, K. (1992). Resampling method under dependent models. *Chin. Ann. Math. B* **13**, 25–34.

Shi, X., Tang, D., and Zhang, Z. (1992). Bootstrap interval estimation for reliability indices from type II censored data (in Chinese). *J. Syst. Sci. Math. Sci.* **12**, 215–220.

Shi, X., and Shao, J. (1988). Resampling estimation when the observations *m*-dependent. *Commun. Statist. Theory Methods* **17**, 3923–3934.

Shi, X., Wu, C. F. J., and Chen, J. (1990). Weak and strong representations for quantile processes from finite populations with application to simulation size in resampling inference. *Can. J. Statist.* **18**, 141–148.

Shimabukuro, F. I., Lazar, S., Dyson, H. B. and Chernick, M. R. (1984). A quasi-optical method for measuring the complex permittivity of materials. *IEEE Trans. Micr. T.* **32**, 659–665.*

Shipley, B. (1996). Exploratory path analysis using the bootstrap with applications in ecology and evolution. *Bull. Ecol. Soc. Am. Suppl. Part 2* **77**, 406.*

Shiue, W.-K., Xu, C.-W. and Rea, C. B. (1993). Bootstrap confidence intervals for simulation outputs. *J. Statist. Comput. Simul.* **45**, 249–255.

Shorack, G. R. (1982). Bootstrapping robust regression. *Commun. Statist. Th. Meth.* **11**, 961–972.*

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.*

Siegel, A. F. (1986). Rarefaction Curves. In *Encyclopedia of Statistical Sciences* **7**, 623–626.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. R. Statist. Soc. B* **43**, 97–99.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.*

Silverman, B. W., and Young, G. A. (1987). The bootstrap: To smooth or not to smooth? *Biometrika* **74**, 469–479.*

Simar, L., and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Manage. Sci.* **44**, 49–61.

Simon. J. L. (1969). *Basic Research Methods in Social Science*. Random House, New York.*

Simon. J. L., and Bruce, P. (1991). Resampling: A tool for everyday statistical work. *Chance* **4**, 22–32.*

Simon. J. L., and Bruce, P. (1995). The new biostatistics of resampling. *M. D. Comput.* **12**, 115–121.*

Simonoff, J. S. (1986). Jackknifing and bootstrapping goodness-of-fit statistics in sparse multinomials. *J. Am. Statist. Assoc.* **81**, 1005–1011.*

Simonoff, J. S. (1994a). Book review of *Computer Intensive Statistical Methods: Validation, Model Selection and Bootstrap* by J.S.U. Hjorth. *J. Am. Statist. Assoc.* **89**, 1559–1560.

Simonoff, J. S. (1994b). Book review of *An Introduction to the Bootstrap* by B. Efron and R. Tibshirani, *J. Am. Statist. Assoc.* **89**, 1559–1560.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer-Verlag, New York.*

Simonoff, J. S., and Reiser, Y. B. (1986). Alternative estimation procedures for $\Pr(X < Y)$ in categorized data. *Biometrics* **42**, 895–907.

Simonoff, J. S., and Tsai, C.-L. (1988). Jackknife and bootstrapping quasi-likelihood. *J. Statist. Comput. Simul.* **30**, 213–232.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9**, 1187–1195.*

Singh, K. (1996). Breakdown theory for bootstrap quantiles. TR # 96–015. Department of Statistics, Rutgers University.

Singh, K. (1997). Book Review of *The Jackknife and Bootstrap* (by J. Shao, and D. Tu), *J. Am. Statist. Assoc.* **92**, 1214.

Singh, K., and Babu, G. J. (1990). On the asymptotic optimality of the bootstrap. *Scand. J. Statist.* **17**, 1–9.

Singh, K., and Liu, R. Y. (1990). On the validity of the jackknife procedure. *Scand. J. Statist.* **17**, 11–21.

Sinha, A. L. (1983). Bootstrap algorithms for parameter and smoothing state estimation. *IEEE Trans. Aerosp. Electron. Syst.* **19**, 85–88.

Sitnikova, T. (1996). Bootstrap method of interior—branch test for phylogenetic trees. *Mol. Biol. Evol.* **13**, 605–611.*

Sitnikova, T., Rzhetsky, A., and Nei, M. (1995). Interior-branch and bootstrap tests of phylogenetic trees. *Mol. Biol. Evol.* **12**, 319–333.*

Sitter, R. R. (1992a). A resampling procedure for complex survey data. *J. Am. Statist. Assoc.* **87**, 755–765.*

Sitter, R. R. (1992b). Comparing three bootstrap methods for survey data. *Can. J. Statist.* **20**, 135–154.*

Sitter, R. R. (1997). Variance estimation for the regression estimator in two-phase sampling. *J. Am. Statist. Assoc.* **92**, 779–787.

Sitter, R. R. (1998). Balanced resampling using orthogonal multiarrays. In *Encyclopedia of Statistical Sciences*, Update Volume 2 (S. Kotz, C. B. Read, and D. L. Banks, editors), pp. 46–50. Wiley, New York.*

Sivaganesan, S. (1994). Book review of *An Introduction to the Bootstrap* (by B. Efron and R. J. Tibshirani). *SIAM Rev.* **36**, 677–678.

Skinner, C. J., Holt, D., and Smith, T. M. F. (editors) (1989). *Analysis of Complex Surveys*. Wiley, New York.

Smith, A. F. M., and Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling–resampling perspective. *J. Am. Statist. Assoc.* **46**, 84–88.

Smith, L. A., and Sielken, R. L. (1988). Bootstrap bounds for "safe" doses in the multistage cancer dose-response model. *Commun. Statist. Simul. Comput.* **17**, 153–175.*

Smith, L. R., Harrell, F. E., and Muhlbaier, L. H. (1992). Problems and potentials in modeling survival. In *Medical Effectiveness Research Data Methods*. (M. L. Grady and H. A. Schwarts, editors) 151–159. U. S. Department of Health and Human Services, Agency for Health Care Policy Research.

Snapinn, S. M., and Knoke, J. D. (1984). Classification error rate estimators evaluated by unconditional mean squared error. *Technometrics* **26**, 371–378.*

Snapinn, S. M., and Knoke, J. D. (1985a). An evaluation of smoothed classification error rate estimators. *Technometrics* **27**, 199–206.*

Snapinn, S. M., and Knoke, J. D. (1985b). Improved classification error rate estimation: Bootstrap or smooth? Unpublished report.*

Snapinn, S. M., and Knoke, J. D. (1988). Bootstrap and smoothed classification error rate estimates. *Commun. Statist. Simul. Comput.* **17**, 1135–1153.*

Solka, J. L., Wegman, E. J., Priebe, C. E., Poston, W. L., and Rogers, G. W. (1995). A method to determine the structure of an unknown mixture using Akaike Information Criterion and the bootstrap. George Mason University Technical Report.

Solomon, H. (1986). Confidence intervals in legal settings. In *Statistics and the Law* (M. H. DeGroot, S. E. Fienberg and J. B. Kadane, editors), pp. 455–473. Wiley, New York.

Solow, A. R. (1985). Bootstrapping correlated data. *J.I.A. Math. Geo.* **17**, 769–775.*

Sorum, M. (1972). Three probabilities of misclassification. *Technometrics* **14**, 309–316.*

Sparks, T. H., and Rothery, P. (1996). Resampling methods for ecotoxicological data. *Ecotoxicology* **5**, 197–207.

Sprent, P. (1998). *Data Driven Statistical Methods*. Chapman & Hall, London.*

Srivastava, M. S. (1987a). Bootstrapping Durbin-Watson statistics. *Indian J. Math.* **29**, 193–210.

Srivastava, M. S. (1987b). Bootstrap methods in ranking and slippage problems. Commun. *Statist. Theory Methods* **16**, 3285–3299.

Srivastava, M. S., and Carter, E. M. (1983). *An Introduction to Applied Multivariate Statistics*. Elsevier Science Publishing Co., New York.*

Srivastava, M. S., and Chan, Y. M. (1989). A comparison of bootstrap methods and Edgeworth expansion in approximation of the distribution of sample variance: One sample and two sample cases. *Commun. Statist. Simul. Comput.* **18**, 339–361.

Srivastava, M. S., and Lee, G. C. (1984). On the distribution of the correlation coefficient when sampling from a mixture of two bivariate normal densities: robustness and outliers. *Can. J. Statist.* **2**, 119–133.*

Srivastava, M. S., and Singh, B. (1989). Bootstrapping in multiplicative models, *J. Econ.* **42**, 287–297.*

Stanghaus, G. (1987). Bootstrap and inference procedures for L_1 regression. In *Statistical Data Analysis Based on the L_1 Norm and Related Methods* (Y. Dodge, editor), pp. 323–332. North-Holland, Amsterdam.

Staudte, R. G., and Sheather, S. J. (1990). *Robust Estimation and Testing*. Wiley, New York.*

Stauffer, D. F., Garton, E. O., and Steinhorst, R. K. (1985). A comparison of principal components from real and random data. *Ecology* **66**, 1693–1698.

Stefanski, L. A., and Cook, J. R. (1995). Simultaneous extrapolation: The measurement error jackknife. *J. Am. Statist. Assoc.* **90**, 1247–1256.

Steinberg, S. M. (1983). Confidence intervals for functions of quantiles using linear combinations of order statistics. Ph.D. dissertation, Department of Statistics, University of North Carolina, Chapel Hill.

Stein, M. (1987). Large sample properties of simulations using Latin hyercube sampling. *Technometrics* **29**, 143–151.*

Stein, M. L. (1989). Asymptotic distributions of minimum norm quadratic estimators of the covariance function of a Gaussian random field. *Ann. Statist.* **17**, 980–1000.

Stewart, T. J. (1986). Experience with a Bayesian bootstrap method incorporating proper prior information. *Commun. Statist. Theory Methods* **15**, 3205–3225.

Stine, R. A. (1982). Prediction intervals for time series. Ph.D. dissertation, Princeton University, Princeton.

Stine, R. A. (1985). Bootstrap prediction intervals for regression. *J. Am. Statist. Assoc.* **80**, 1026–1031.*

Stine, R. A. (1987). Estimating properties of autoregressive forecasts. *J. Am. Statist. Assoc.* **82**, 1072–1078.*

Stine, R. A. (1992). An introduction to bootstrap methods. *Sociol. Methods Res.* **18**, 243–291.*

Stine, R. A., and Bollen, K. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In *Testing Structural Equation Models*. Sage Publications, Beverly Hills.

Stoffer, D. S., and Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and Kalman filter. *J. Am. Statist. Assoc.* **86**, 1024–1033.*

Strawderman, R. L., Parzen, M. I., and Wells, M. T. (1997). Accurate confidence limits for quantiles under random censoring. *Biometrics* **53**, 1399–1415.

Strawderman, R. L., and Wells, M. T. (1997). Accurate bootstrap confidence limits for the cumulative hazard and survivor functions under random censoring. *J. Am. Statist. Assoc.* **92**, 1356–1374.

Stromberg, A. J. (1997). Robust covariance estimates based on resampling. *J. Statist. Plann. Inf.* **57**, 321–334.

Stuart, A., and Ord, K. (1993). *Kendall's Advanced Theory of Statistics,* Vol. 1, 6th ed., Edward Arnold, London.*

Stute, W. (1990). Bootstrap of the linear correlation model. *Statistics* **21**, 433–436.

Stute, W. (1992). Modified cross-validation in density estimation. *J. Statist. Plann. Inf.* **30**, 293–305.

Stute, W., and Grunder, B. (1992). Bootstrap approximations to prediction intervals for explosive AR(1)-processes. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 121–130. Springer-Verlag, Berlin.

Stute, W., and Wang, J. (1994). The jackknife estimate of a Kaplan–Meier integral. *Biometrika* **81**, 602–606.

Stute, W., Manteiga, W. C., and Quindimil, M. P. (1993). Bootstrap based goodness-of-fit tests. *Metrika* **40**, 243–256.

Sun, L., and Muller-Schwarze, D. (1996). Statistical resampling methods in biology: A case study of beaver dispersal patterns. *Am. J. Math. Manage. Sci.* **16**, 463–502.*

Sutton, C. D. (1993). Computer-intensive methods for tests about the mean of an asymmetric distribution. *J. Am. Statist. Assoc.* **88**, 802–810.

Swanepoel, J. W. H. (1983). Bootstrap selection procedures based on robust estimators. *Commun. Statist. Theory Methods* **12**, 2059–2083.

Swanepoel, J. W. H. (1985). Bootstrap selection procedures based on robust estimators. In *The Frontiers of Modern Statistical Inference Procedures* (E. J. Dudewicz, editor), pp. 45–64. American Science Press, Columbus.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works. *Commun. Statist. Theory Methods* **15**, 1399–1415.

Swanepoel, J. W. H., and van Wyk, J. W. J. (1986). The bootstrap applied to power spectral density function estimation. *Biometrika* **73**, 135–141.*

Swanepoel, J. W. H., van Wyk, J. W. J., and Venter, J. H. (1983). Fixed width confidence intervals based on bootstrap procedures. *Seq. Anal.* **2**, 289–310.

Swift, M. B. (1995). Simple confidence intervals for standardized rates based on the approximate bootstrap method. *Statist. Med.* **14**, 1875–1888.

Takeuchi, L. R., Sharp, P. A., and Ming-Tung, L. (1994). Identifying data generating families of probability densities: A bootstrap resampling approach. In *1994 Proceedings Decision Sciences*, Vol. 2, pp. 1391–1393.

Tambour, M., and Zethraeus, N. (1998). Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results. *Health Econ.* **7**, 143–147.*

Tamura, H., and Frost, P. A. (1986). Tightening CAV (DUS) bounds by using a parametric model. *J. Acct. Res.* **24**, 364–371.

Taylor, C. C. (1989). Bootstrap choice of smoothing parameter in kernel density estimation. *Biometrika* **76**, 705–712.

Taylor, M. S., and Thompson, J. R. (1992). A nonparametric density estimation based resampling algorithm. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 399–403. Wiley, New York.*

ter Braak, C. J. F. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In *Bootstrapping and Related Techniques. Proceedings, Trier, FRG. Lecture Notes in Economics and Mathematical Systems* (K.-H. Jockel, G. Rothe, and W. Sendler, editors), Vol. 376, pp. 70–85. Springer-Verlag, Berlin.

Theodossiou, P. T. (1993). Predicting shifts in the mean of a multivariate time series process: An application in predicting business failures. *J. Am. Statist. Assoc.* **88**, 441–449.

Therneau, T. (1983). Variance reduction techniques for the bootstrap. Ph.D. dissertation, Department of Statistics, Stanford University, Stanford.*

Thisted, R. A. (1988). *Elements of Statistical Computing: Numerical Computation*. Chapman & Hall, New York.*

Thomas, G. E. (1994). Book review of *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (by P. H. Westfall and S. S. Young). *Statistician* **43**, 467–486.

Thombs, L. A., and Schucany, W. R. (1990). Bootstrap prediction intervals for autoregression. *J. Am. Statist. Assoc.* **85**, 486–492.*

Thompson, J. R. (1989). *Empirical Model Building*. Wiley, New York.*

Tibshirani, R. (1985). Bootstrap computations. *SAS SUGI* **10**, 1059–1063.

Tibshirani, R. (1986). Bootstrap confidence intervals. *Proc. Comp. Sci. Statist.* **18**, 267–273.

Tibshirani, R. (1988). Variance stabilization and the bootstrap. *Biometrika* **75**, 433–444.

Tibshirani, R. (1992). Some applications of the bootstrap in complex problems. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 271–277. Wiley, New York.*

Tibshirani, R. (1996). Regression shrinkage and selection via lasso. *J. R. Statist. Soc. B* **58**, 267–288.

Tibshirani, R. (1997a). Who is the fastest man in the world? *Am. Statist.* **51**, 106–111.

Tibshirani, R. (1997b). The lasso method for variable selection in the Cox model. *Statist. Med.* **16**, 385–395.

Tingley, M., and Field, C. (1990). Small-sample confidence intervals. *J. Am. Statist. Assoc.* **85**, 427–434.*

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley, Chichester.*

Tomasson, H. (1995). Risk scores from logistic regression: Unbiased estimates of relative and attributable risk, *Statist. Med.* **14**, 1331–1339.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis, Lecture Notes in Statistics, Vol. 21*. Springer-Verlag, New York.*

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Clarendon Press, Oxford.*

Tourassi, G. D., Floyd, C. E., Sostman, H. D., and Coleman, R. E. (1995). Performanc evaluation of an artificial neural network for the diagnosis of acute pulmonary embolism using the cross-validation, jackknife, and bootstrap methods. *WCNN '95* **2**, 897–900.

Toussaint, G. T. (1974). Bibliography on estimation on misclassification. *IEEE Trans. Inform. Theory* **20**, 472–479.*

Tran, Z. V. (1996). The bootstrap: verifying mean changes in cholesterol from resistive *exercise*. *Med. Sci. Sports Exerc. Suppl.* **28**, 187.*

Tivang, J. G., Nienhuis, J., and Smith, O. S. (1994). Estimation of sampling variance of molecular marker data using the bootstrap procedure. *Theor. Appl. Genetics* **89**, 259–264.*

Troendle, J. F. (1995). A stepwise resampling method of multiple hypothesis testing. *J. Am. Statist. Assoc.* **90**, 370–378.

Tsay, R. S. (1992). Model checking via parametric bootstraps in time series. *Appl. Statist.* **41**, 1–15.*

Tsodikov, A., Hasenclever, D., and Loeffler, M. (1998). Regression with bounded outcome score: Estimation of power by bootstrap and simulation in a chronic myelogenous leukemia clinical trial. *Statist. Med.* **17**, 1909–1922.*

Tsumoto, S., and Tanaka, H. (1995). Automated selection of rule induced methods based on recursive iteration of resampling methods and multiple statistical testing. In *1st International Conference on Knowledge Discovery and Data Mining* (U. M. Fayyad, and R. Uthurusamy, editors), Vol. 1, pp. 312–317. AAAI Press.

Tsumoto, S., and Tanaka, H. (1996a). Automated acquisition of medical expert system rules based on rough sets and resampling methods. In *Proceedings of the Third World Congress on Expert Systems* (J. K. Lee, J. Liebowitz, and Y. M. Chae, editors), Vol. 2, pp. 877–884.

Tsumoto, S., and Tanaka, H. (1996b). Induction of expert system rules from databases based on rough set theory and resampling methods. In *Foundations of Intelligent Systems, 9th International Symposium ISMIS '96 Proceedings* (Z. W. Ras and M. Michalewicz, editors), pp. 128–138. Springer-Verlag, Berlin.

Tu. D. (1986). Bootstrapping of *L*-statistics. *Kexue Tongbao (Chin. Bull. Sci.)* **31**, 965–969.

Tu, D. (1988a). The kernel estimator of conditional *L*-functional and its bootstrapping statistics. *Acta Math. Appl. Sin.* **11**, 53–68.

Tu, D. (1988b). The nearest neighbor estimate of the conditional *L*-functional and its bootstrapping statistics. *Chin. Ann. Math. A* **8**, 345–357.

Tu, D. (1989). *L*-functional and nonparametric *L*-regression estimates: Asymptotic distributions and bootstrapping approximations. Technical Report 89-51, Center for Multivariate Analysis, Pennsylvania State University, State College.

Tu, D. (1992). Approximating the distribution of a general standardized functional statistic with that of jackknife pseudovalues. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, editors), pp. 279–306. Wiley, New York.*

Tu, D., and Cheng, P. (1989). Bootstrapping untrimmed *L*-statistics. *J. Sys. Sci. Math. Sci.* **9**, 14–23.

Tu, D., and Gross, A. J. (1994). Bias reduction for jackknife skewness estimators. *Commun. Statist. Theory Methods* **23**, 2323–2341.

Tu, D., and Gross, A. J. (1995). Accurate confidence intervals for the ratio of specific occurrence/exposure rates in risk and survival analysis. *Biometric J.* **37**, 611–626.

Tu, D., and Shi, X. (1988). Bootstrapping and randomly weighting the *U*-statistics with jackknifed pseudo values. *Math. Statist. Appl. Probab.* **3**, 205–212.

Tu, D., and Zhang, L. (1992a). On the estimation of skewness of a statistic using the jackknife and the bootstrap. *Statist. Hefte* **33**, 39–56.

Tu, D., and Zhang, L. (1992b). Jackknife approximations for some nonparametric confidence intervals of functional parameters based on normalizing transformations. *Comput. Statist.* **7**, 3–15.

Tu, D., and Zheng, Z. (1987). On the Edgeworth's expansion of random weighting method. *Chin. J. Appl. Probab. Statist.* **3**, 340–347.

Tu, D., and Zheng, Z. (1991). Random weighting: Another approach to approximate the unknown distributions of pivotal quantities. *J. Comb. Inf. Syst. Sci.* **16**, 249–270.

Tu, X. M., Burdick, D. S., and Mitchell, B. C. (1992). Nonparametric rank estimation using bootstrap resampling and canonical correlation analysis. In *Exploring the Limits of Bootstrap* (R. LePage, and L. Billard, editors), pp. 405–418. Wiley, New York.*

Tucker, H. G. (1959). A generalization of the Glivenko–Cantelli theorem. *Ann. Math. Statist.* **30**, 828–830.*

Tukey, J. W. (1958). Bias and confidence in not quite large samples (abstract). *Ann. Math. Statist.* **29**, 614.*

Turnbull, B. W., and Mitchell, T. J. (1978). Exploratory analysis of disease prevalence data from sacrifice experiments. *Biometrics* **34**, 555–570.*

Turner, S., Myklebust, R. L., Thorne, B. B., Leigh, S. D., and Steel, E. B. (1996). Airborne asbestos method; bootstrap method for determining the uncertainty of asbestos concentration. National Institute of Standards and Technology Technical Report.

Ueda, N., and Nakano, R. (1995). Estimating expected error rates of neural network classifiers in small sample size situations: A comparison of cross-validation and bootstrap. In *1995 IEEE International Conference on Neural Networks, Proceedings* **1**, 101–104.*

Upton, G. J. G., and Fingleton, B. (1985). *Spatial Data Analysis by Example,* Vol. 1*: Point Pattern and Quantitative Data*. Wiley, Chichester.*

Upton, G. J. G., and Fingleton, B. (1989). *Spatial Data Analysis by Example,* Vol. 2*: Categorical and Directional Data*. Wiley, Chichester.*

van der Burg, E., and de Leeuw, J. (1988). Use of the multinomial, jackknife and bootstrap in generalized nonlinear canonical correlation analysis. *Appl. Stoc. Models Data Anal.* **4**, 159–172.

van der Kloot, W. (1996). Statistics for studying quanta at synapses: Resampling and confidence limits on histograms. *J. Neurosci. Methods* **65**, 151–155.

van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag, New York.*

van Dongen, S. (1995). How should we bootstrap allozyme data? *Heredity* **74**, 445–447.

van Dongen, S., and Backeljau, T. (1995). One- and two-sample tests for single-locus inbreeding coefficients using the bootstrap. *Heredity* **74**, 129–135.

van Dongen, S., and Backeljau, T. (1997). Bootstrap tests for specific hypotheses at single-locus inbreeding coefficients. *Genetica* **99**, 47–58.

van Zwet, W. (1989). Hoeffding's decomposition and the bootstrap. Talk given at Conference on Asymptotic methods for computer-intensive procedures in statistics. Oberwolfach, Germany.

Veall, M. R. (1989). Applications of computationally-intensive methods to econometrics. In *Proceedings of the 47th Session of the International Statistical Institute, Paris*, pp. 75–88.

Ventura, V. (1997). Likelihood inference by Monte Carlo methods and efficient nested bootstrapping. D.Phil. Thesis, Department of Statistics, Oxford University.*

Ventura, V., Davison, A. C., and Boniface,S. J. (1997). Statistical inference for the effect of magnetic brain stimulation on a motoneurone. *Appl. Statist.* **46**, 77–94.*

Vinod, H. D., and McCullough, B. D. (1995). Estimating cointegration parameters: an application of the double bootstrap. *J. Statist. Plan. Inf.* **43**, 147–156.

Vinod, H. D., and Raj, B. (1988). Econometric issues in Bell System diverstiture: A bootstrap application. *Appl. Statist.* **37**, 251–261.

Visscher, P. M., Thompson, R., and Haley, C. S. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013–1020.

Wacholder, S., Gail, M. H., Pee, D., and Brookmeyer, R. (1989). Alternative variance and efficiency calculations for the case-cohort design. *Biometrika* **76**, 117–123.

Waclawiw, M. A., and Liang, K. (1994). Empirical Bayes estimation and inference for random effects model with binary response. *Staist. Med.* **13**, 541–551.*

Wagner, R. F., Chan, H.-P., Sahiner, B., Petrick, N., and Mossoba, J. T. (1997). Finite-sample effects and resampling plans: Applications to linear classifiers I computer-aided diagnosis. *Proc. SPIE* **3034**, 467–477.

Wahrendorf, J., Becher, H., and Brown, C. C. (1987). Bootstrap comparison of non-nested generalized linear models: Applications in survival analysis and epidemiology. *Appl. Statist.* **36**, 72–81.

Wahrendorf, J., and Brown, C. C. (1980). Bootstrapping a basic inequality in the analysis of joint action of two drugs. *Biometrics* **36**, 653–657.*

Wallach, D., and Goffinet, B. (1987). Mean squared error of prediction in models for studying ecological and agronomic systems. *Biometrics* **43**, 561–573.

Walther, G. (1997). Granulometric smoothing. *Ann. Statist.* **25**, 2273–2299.

Wang, M.-C. (1986). Re-sampling procedures for reducing bias of error rate estimation in multinomial classification. *Comput. Statist. Data Anal.* **4**, 15–39.

Wang, J.-L., and Hettmansperger, T. P. (1990). Two-sample inference for median survival times based on one-sample procedures for censored survival data. *J. Am. Statist. Assoc.* **85**, 529–536.

Wang, S. J. (1989). On the bootstrap and smoothed bootstrap. *Commun. Statist. Theory Methods* **18**, 3949–3962.

Wang, S. J. (1990). Saddlepoint approximations in resampling analysis. *Ann. Inst. Statist. Math.* **42**, 115–131.

Wang, S. J. (1992). General saddlepoint approximations in the bootstrap. *Statist. Probab. Lett.* **13**, 61–66.

Wang, S. J. (1993a). Saddlepoint expansions in finite population problems. *Biometrika* **80**, 583–590.

Wang, S. J. (1993b). Saddlepoint methods for bootstrap confidence bands in nonparametric regression. *Aust. J. Statist.* **35**, 93–101.

Wang, S. J. (1995). Optimizing the smoothed bootstrap. *Ann. Inst. Statist. Math.* **47**, 65–80.

Wang, S. J., Woodward, W. A., Gray, H. L., Wiechecki, S., and Sain, S. R. (1997). A new test for outlier detection from a multivariate mixture distribution. *J. Comp. Graph. Statist.* **6**, 285–299.

Wang, X., and Mong, J. (1994). Resampling-based estimator in nonlinear regression. *Statist. Sin.* **4**, 187–198.

Wang, Y. (1996). A likelihood ratio test against stochastic ordering in several populations. *J. Am. Statist. Assoc.* **91**, 1676–1683.

Wang, Y., Prade, R. A., Griffith, J. Timberlake, W. E., and Arnold, J. (1994). Assessing the statistical reliability of physical maps by bootstrap resampling. *Comput. Appl. Biosci.* **10**, 625–634,

Wang, Y., and Wahba, G. (1995). Bootstrap confidence intervals for smoothing splines and their comparison to Bayesian confidence intervals. *J. Statist. Comput. Simul.* **51**, 263–279.*

Ware, J. H., and De Gruttola, V. (1985). Multivariate linear models for longitudinal data: A bootstrap study of the GLS estimate. In *Biostatistics in Biomedical, Public Health and Environmental Sciences* (P. K. Sen, editor), pp. 424–434. North-Holland, Amsterdam.

Wasserman, G. S., Mohsen, H. A., and Franklin, L. A. (1991). A program to calculate bootstrap confidence intervals for process capability index, C_{pk}. *Commun. Statist. Simul. Comput.* **20**, 497–510.*

Watson, G. S. (1983). The computer simulation treatment of directional data. In *Proceedings of the Geological Conference, Kharagpur India*, India, *Indian J. Earth Sci.* 19–23.

Weber, N. C. (1984). On resampling techniques for regression models. *Statist. Probab. Lett.* **2**, 275–278.*

Weber, N. C. (1986). On the jackknife and bootstrap techniques for regression models. In *Pacific Statistical Congress: Proceedings of the Congress* (I. S. Francis, B. F. J. Manly, and F. C. Lam, editors), pp. 51–55. North-Holland, Amsterdam.

Weinberg, S. L., Carroll, J. D., and Cohen, H. S. (1984). Confidence regions for IDSCAL using the jackknife and bootstrap techniques. *Psychometrika* **49**, 475–491.

Weiss, G. (1975). Time-reversibility of linear stochastic processes. *J. Appl. Probab.* **12**, 831–836.*

Weiss, I. M. (1970). A survey of discrete Kalman-Bucy filtering with unknown noise covariances. AIAA Guidance, Control and Flight Mechanics Conference, AIAA Paper No. 70-955, American Institute of Aeronautics and Astronautics, New York.*

Weissfeld, L. A., and Schneider, H. (1987). Inference based on the Buckley–James procedure. *Commun. Statist. Theory Methods* **16**, 1773–1787.

Welch, B. L., and Peers, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Statist. Soc. B* **25**, 318–329.*

Welch, W. J. (1990). Construction of permutation tests. *J. Am. Statist. Assoc.* **85**, 693–698.

Wellner, J. A., and Zhan,Y. (1996). Bootstrapping z-estimators. Techical Report, University of Washington, Department of Statistics.

Wellner, J. A., and Zhan,Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J. Am. Statist. Assoc.* **92**, 945–959.

Wells, M. T., and Tiwari, R. C. (1994). Bootstrapping a Bayes estimator of a survival function with censored data. *Ann. Inst. Statist. Math.* **46**, 487–495.

Wendel, M. (1989). Eine anwendung der bootstrap—Methode auf nicht-lineare autoregressive prozesse erster ordnung. Diploma thesis, Technical University of Berlin, Berlin.

Weng, C.-S. (1989). On a second order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.* **17**, 705–710.*

Wernecke, K.-D., and Kalb, G. (1987). Estimation of error rates by means of simulated bootstrap distributions. *Biom. J.* **29**, 287–292.

Westfall, P. (1985). Simultaneous small-sample multivariate Bernoulli confidence intervals. *Biometrics* **41**, 1001–1013.*

Westfall, P., and Young, S. S. (1989). *P* value adjustments for multiple tests in multivariate binomial models. *J. Am. Statist. Assoc.* **84**, 780–786.

Westfall, P., and Young, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley, New York.*

Willemain, T. R. (1994). Bootstrapping on a shoestring: resampling using spreadsheets. *Am. Statist.* **48**, 40–42.

Withers, C. S. (1983). Expansion for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Ann. Statist.* **11**, 577–587.

Wolfe, J. H. (1967). NORMIX: Computational methods for estimating the parameters of multivariate normal mixtures of distributions. Research Memo. SRM 68-2. U.S. Naval Personnel Research Activity, San Diego.*

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivar. Behav. Res.* **5**, 329–350.*

Wong, M. A. (1985). A bootstrap testing procedure for investigating the number of subpopulations. *J. Statist. Comput. Simul.* **22**, 99–112.

Woodroof, J. B. (1997). Statistically comparing user satisfaction instruments: An application of the bootstrap using a spreadsheet. *Proceedings of the 13th Hawaii Internaional Conference on System Sciences*, Vol. 3, pp. 49–56.

Woodroofe, M., and Jhun, M. (1989). Singh's theorem in the lattice case. *Statist. Probab. Lett.* **7**, 201–205.

Worton, B. J. (1994). Book review of *Computer Intensive Statistical Methods: Validation, Model Selection and Bootstrap* by J. S. U. Hjorth. *J. R. Statist. Soc. A* **157**, 504–505.

Worton, B. J. (1995). Modeling radio-tracking data. *Environ. Ecol. Statist.* **2**, 15–23.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling plans in regression analysis (with discussion). *Ann. Statist.* **14**, 1261–1350.*

Wu, C. F. J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18**, 1438–1452.

Wu, C. F. J. (1991). Balanced repeated replications based on mixed orthogonal arrays. *Biometrika* **78**, 181–188.

Xie, F., and Paik, M. C. (1997). Multiple imputation methods for the missing covariates in generalized estimating equations. *Biometrics* **53**, 1538–1546.

Yandell, B. S., and Horvath, L. (1988). Bootstrapped multi-dimensional product limit process. *Aust. J. Statist.* **30**, 342–358.

Yang, S. S. (1985a). On bootstrapping a class of differentiable statistical functionals with application to *L*- and *M*-estimates. *Statist. Neerl.* **39**, 375–385.

Yang, S. S. (1985b). A smooth nonparametric estimator of a quantile function. *J. Am. Statist. Assoc.* **80**, 1004–1011.

Yang, S. S. (1988). A central limit theorem for the bootstrap mean. *Am. Statist.* **42**, 202–203.*

Yang, Z. R., Zwolinski, M., and Chalk, C. D. (1998). Bootstrap, an alternative to Monte Carlo simulation. *Electron. Lett.* **34**, 1174–1175.

Yeh, A. B., and Singh, K. (1997). Balanced confidence regions based on Tukey's depth and the bootstrap. *J. R. Statist. Soc. B* **59**, 639–652.

Yokoyama, S., Seki, T., and Takashina, T. (1993). Bootstrap for the normal parameters. *Commun. Statist. Simul. Comput.* **22**, 191–203.

Young, G. A. (1986). Conditioned data-based simulations: Some examples from geometrical statistics. *Int. Statist. Rev.* **54**, 1–13.

Young, G. A. (1988a). A note on bootstrapping the correlation coefficient. *Biometrika* **75**, 370–373.*

Young, G. A. (1988b). Resampling tests of statistical hypotheses. *Proceedings of the Eighth Biannual Symposium on Computational Statistics* (D. Edwards and N. E. Raum, editors), pp. 233–238. Physica-Verlag, Heidelberg.

Young, G. A. (1990). Alternative smoothed bootstraps. *J. R. Statist. Soc. B* **52**, 477–484.

Young, G. A. (1993). Book review of *The Bootstrap and Edgeworth Expansion* (by P. Hall). *J. R. Statist. Soc. A* **156**, 504–505.

Young, G. A. (1994). Bootstrap: More than a stab in the dark? (with discussion). *Statist. Sci.* **9**, 382–415.*

Young, G. A., and Daniels, H. E. (1990). Bootstrap bias. *Biometrika* **77**, 179–185.*

Youyi, C., and Tu, D. (1987). Estimating the error rate in discriminant analysis: By the delta, jackknife and bootstrap method (in Chinese). *Chin. J. Appl. Probab. Statist.* **3**, 203–210.

Yu, Z., and Tu, D. (1987). On the convergence rate of bootstrapped and randomly weighted *m*-dependent means. Research Report, Institute of Systems Science, Academia Sinica, Beijing.

Yuen, K. C., and Burke, H. D. (1997). A test of fit for a semiparametric additive risk model. *Biometrika* **84**, 631–639.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's Sunspot Numbers. *Philos. Trans. A* **226**, 267–298.*

Zecchi, S., and Camillo, F. (1994). An application of bootstrap to the Rv coefficient in the analysis of stock markets (Italian). *Quand. Statist. Mat, Appl.* **14**, 99–109.

Zelterman, D. (1993). A semiparametric bootstrap technique for simulating extreme order statistics. *J. Am. Statist. Assoc.* **88**, 477–485.*

Zeng, Q., and Davidian, M. (1997a). Bootstrap-adjusted calibration confidence intervals for immunoassay. *J. Am. Statist. Assoc.* **92**, 278–290.

Zeng, Q., and Davidian, M. (1997b). Testing homogeneity of intra-run variance parameters in immunoassay. *Statist. Med.* **16**, 1765–1776.

Zhan, Y. (1996). Bootstrapping functional *M*-estimators. Unpublished Ph.D. dissertation. University of Washington, Department of Statistics.

Zhang, J., and Boos, D. D. (1992). Bootstrap critical values for testing homogeneity of covariance matrices. *J. Am. Statist. Assoc.* **87**, 425–429.

Zhang, J., and Boos, D. D. (1993). Testing hypotheses about covariance matrices using bootstrap methods. *Commun. Statist. Theory Methods* **22**, 723–739.

Zhang, L., and Tu, D. (1990). A comparison of some jackknife and bootstrap procedures in estimating sampling distributions of studentized statistics and constructing confidence intervals. Tech. Report 90-28, Center for Multivariate Analysis, Pennsylvania State University, State College.

Zhang, Y, Hatzinakos, D., and Venetsanopoulos, A. N. (1993). Bootstrapping techniques in the estimation of higher-order cumulants from short data records. *Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 4, pp. 200–203.

Zhao, L. (1986). Bootstrapping the error variance estimate in linear model (in Chinese). *A. Ma. Peking* **29**, 36–45.

Zharkikh, A., and Li W.-H. (1992). Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences: II. Four taxa without a molecular clock. *J. Mol. Evol.* **4**, 44–63.*

Zharkikh, A., and Li W.-H. (1995). Estimation of confidence in phylogeny: Complete-and-partial bootstrap technique. *Mol. Phylogenet. Evol.* **4**, 44–63.*

Zheng, X. (1994). Third-order correct bootstrap calibrated confidence bounds for nonparametric mean. *Math. Methods Statist.* **3**, 62–75.

Zheng, Z. (1985). The asymptotic behavior of the nearest neighbour estimate and its bootstrap statistics. *Sci. Sin. A.* **28**, 479–494.

Zhou, M. (1993). Bootstrapping the survival curve estimator when data are doubly censored. Technical Report 335. University of Kentucky, Department of Statistics.

Zhu, L. X., and Fang, K. T. (1994). The accurate distribution of the Kolmogorov statistic with applications to bootstrap approximation. *Adv. Appl. Math.* **15**, 476–489.

Zhu, W. (1997). Making bootstrap statistical inferences: A tutorial. *Res.Q. Exerc. Sport.* **68**, 44–55.

Ziari, H. A., Leatham, D. J., and Ellinger, P. N. (1997). Developments of statistical discriminant mathematical programming model via resampling estimation techniques. *Am. J. Agric. Econ*, **79**, 1352–1362.

Ziegel, E. (1994). Book review of *Bootstrapping: A Nonparametric Approach to Statistical Inference* (by C. Z. Mooney and R. Duvall). *Technometrics* **36**, 435–436.

Ziliak, J. P. (1997). Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators, *J. Bus. Econ. Statist.* **15**, 419–431.

Zoubir, A. M. (1994a). Bootstrap multiple tests: An application of to optimal sensor location for knock detection. *Appl. Signal Process.* **1**, 120–130.

Zoubir, A. M. (1994b). Multiplr bootstrap testsand their application. *1994 IEEE International Conference on Acoustics, Speech, and Signal Processing* **6**, 69–72.

Zoubir, A. M., and Ameer, I. (1997). Bootstrap analysis of polynomial amplitude and phase signals. In *Proceedings of IEEE TENCON '97, IEEE Region 10 Annual Conference, Speech and Image Technologies for Computing and Telecommunications* (M. Deriche, M. Moody, and M. Bennamoun, editors), Vol. 2, pp. 843–846.

Zoubir, A. M., and Boashash, B. (1998). The bootstrap and its application in signal processing. *IEEE Signal Process. Mag.* **15**, 56–76.*

Zoubir, A. M., and Bohme, J. F. (1995). Bootstrap multiple tests applied to sensor location. *IEEE Trans. Signal Process.* **43**, 1386–1396.*

Zoubir, A. M., and Iskander, D. R. (1996). Bispectrum based Gaussianity test using the bootstrap. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. **5**, 3029–3032.

Zoubir, A. M., Iskander, D. R., Ristac, B., and Boashash, B. (1994). Instantaneous frequency estimation: Confidence bounds using the bootstrap. In *Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers* (A. Singh, editor), Vol. 2, pp. 55–59.

Bibliography 2 (1999–2007)

1999

Abdelhafez, M. E. M. (1999). Statistical inference for generalized log Burr distribution. *Egyptian Statist. J.* **43**, 209–222.

Aerts, M., and Claeskens, G. (1999). Bootstrapping pseudolikelihood models for clustered binary data. *Ann. Inst. Statist. Math.* **51**, 515–530.

Allen, M., and Datta, S. (1999). A note on bootstrapping *M*-estimators in ARMA models. *J. Time Series Anal.* **20**, 365–379.

Aronsson, M., Arvastson, L., Holst, J., Lindoff, B., and Svensson, A. (1999). Bootstrap control. In *Proceedings of the American Control Conference*, pp. 4516–4520. American Automatic Control Council, Evanston, IL.

Athreya, K. B., Fukuchi, J.-I., and Lahiri, S. N. (1999). On the bootstrap and the moving block bootstrap for the maximum of a stationary process. *J. Statist. Plann. Inf.* **76**, 1–17.

Babu, G. J. (1999). Breakdown theory for estimators based on bootstrap and other resampling schemes. In *Asymptotics, Nonparametrics, and Time Series—A Tribute to Madan Lal Puri* (S. Ghosh, editor), 669–681. Marcel Dekker, New York.*

Babu, G. J., Padmanabhan, A. R., and Puri, M. L. (1999). Robust one-way ANOVA under possibly non-regular conditions. *Biomed. J.* **41**, 321–339.

Babu, G. J., Pathak, P. K., and Rao, C. R. (1999). Second-order correctness of the Poisson bootstrap. *Ann. Statist.* **27**, 1666–1683.

Bakke, P. D., Thomas, R., and Parrett, C. (1999). Estimation of long-term discharge statistics by regional adjustment. *J. Am. Water Res. Assoc.* **35**, 911–921.

Barber, S., and Jennison, C. (1999). Symmetric tests and confidence intervals for survival probabilities and quantiles of censored survival data. *Biometrics* **55**, 430–436.

Bertail, P., Politis, D. N., and Romano, J. P. (1999). On subsampling estimators with unknown rate of convergence. *J. Am. Statist. Assoc.* **94**, 569–579.

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

Bickel, P. J., and Bühlmann, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5**, 413–446.

Bingham, N. H., and Pitts, S. M. (1999). Nonparametric inference from M/G/1 busy periods. *Commun. Statist. Stoch. Models* **15**, 247–272.

Bolle, R. M., Ratha, N. K., and Pankanti, S. (1999). Evaluating authentication systems using bootstrap confidence intervals. In *Proceedings of AutoID*, pp. 9–13, Summit, NJ.

Booth, J. G., Hobert, J. P., and Ohman, P. A. (1999). On the probable error of the ratio of two gamma means. *Biometrika* **86**, 439–452.

Brodeen, A. E. M., Brundick, F. S., and Taylor, M. S. (1999). A statistical approach to corpus generation. In *ASA Proceedings of the Section on Section on Physical and Engineering Sciences*, pp. 97–100. American Statistical Association, Alexandria, VA.

Bühlmann, P., and Künsch, H. R. (1999a). Block length selection in the bootstrap for time series. *Comput. Statist. Data Anal.* **31**, 295.

Bühlmann, P., and Künsch, H. R. (1999b). Comments on "Prediction of spatial cumulative distribution functions using subsampling." *J. Am. Statist. Assoc.* **94**, 97–99.

Bühlmann, P., and Wyner, A. J. (1999). Variable length Markov chains. *Ann. Statist.* **27**, 480–513.

Carpenter, J. (1999). Test inversion bootstrap confidence intervals. *J. R. Statist. Soc. B* **61**, 159–172.

Cerf, R., and Cirillo, E. N. M. (1999). Finite size scaling in three-dimensional bootstrap percolation. *Ann. Probab.* **27**, 1837–1850.

Chakraborty, B. (1999). On multivariate median regression. *Bernoulli* **5**, 683–703.

Chan, W., Leung, K., Chan, D. K.-S., Ho, R. M., and Yung, Y.-F. (1999). An alternative method for evaluating congruence coefficients with procrustes rotation: A bootstrap procedure. *Psych. Methods* **4**, 378–402.

Chen, H., and Romano, J. P. (1999). Bootstrap-assisted goodness-of-fit tests in the frequency domain. *J. Time Series Anal.* **20**, 619–654.

Chen, L., and Shao, J. (1999). Bootstrap minimum cost estimation of the average chemical concentration in contaminated soils. *EnvironMetrics* **10**, 153–161.

Chen, X., and Fan, Y. (1999). Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *J. Econ.* **91**, 373–401.

Cheng, M.-Y., and Hall, P. (1999). Mode testing in difficult cases. *Ann. Statist.* **27**, 1294–1315.

Chen-Mok, M., and Sen, P. K. (1999). Nondifferentiable errors in beta-compliance integrated logistic models. *Commun. Statist. Th. Meth.* **28**, 931–946.

Chernick, M. R. (1999a). Bootstrap confidence intervals for C_{pk}: An application to catheter lesion data. In *ASA Proceedings of the Section on Quality and Productivity*, pp. 46–49. American Statistical Association, Alexandria, VA.*

Chernick, M. R. (1999b). *Bootstrap Methods: A Practitioner's Guide*. Wiley, New York.*

Christoffersson, J. (1999). Resampling a nonlinear regression model in the frequency domain. *Commun. Statist. Simul. Comput.* **28**, 329–348.

Ciarlini, P., Gigli, A., and Regoliosi, G. (1999). The computation of accuracy of quality parameters by means of a Monte Carlo simulation. *Commun. Statist. Simul. Comput.* **28**, 821–848.

Cohn, R. D. (1999). Comparisons of multivariate relational structures in serially correlated data. *J. Agr. Biol. Environ. Statist.* **4**, 238–257.

Cole, S. R. (1999). Simple bootstrap statistical inference using the SAS system. *Comput. Meth. Prog. Biomed.* **60**, 79–82.

Commandeur, J. J. F., Groenen, P. J. F., and Meulman, J. J. (1999). A distance-based variety of nonlinear multivariate data analysis, including weights for objects and variables. *Psychometrika* **64**, 169–186.

Cribari-Neto, F., and Zarkos, S. G. (1999a). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. In *Proceedings of the 12th Conference of the Greek Statistical Institute*, pp. 327–339. Greek Statistical Institute, Greece.

Cribari-Neto, F., and Zarkos, S. G. (1999b). Bootstrap methods for heteroskedastic regression models: Evidence on estimation and testing. *Econ. Rev.* **18**, 211–228.

Das, S., and Krishen, A. (1999). Some bootstrap methods in nonlinear mixed-effect models. *J. Statist. Plann. Inf.* **75**, 237–245.

Davidson, R., and MacKinnon, J. G. (1999). The size distortion of bootstrap tests. *Econ. Theory* **15**, 361–376.

DePatta, P. V. (1999). How sharp are classifications? *Ecology* **80**, 2508–2516.

del Barrio, E., Matran, C., Matrán, C., and Cuesta-Albertos, J. A. (1999). Necessary conditions for the bootstrap of the mean of a triangular array. *Ann. Inst. Henri Poincaré: Probab. Statist.* **35**, 371–386.

de Menezes, L. M. (1999). On fitting latent class models for binary data: The estimation of standard errors. *Brit. J. Math. Statist. Psych.* **52**, 149–168.

Di Battista, T., and Di Spalatro, D. (1999). A bootstrap method for adaptive cluster sampling. *Classification and Data Analysis: Theory and Application—Proceedings of the Biannual Meeting of the Classification Group of the Società Italiana di Statistica (SIS)* (M. Vichi, and O. Opitz, editors), pp. 19–26. Springer-Verlag, Berlin.

Dielman, T. E., and Rose, E. L. (1999). Bootstrap versus traditional hypothesis testing procedures for coefficients in least absolute value regression. In *ASA Proceedings of the Business and Economic Statistics Section*, pp. 245–250. American Statistical Association, Alexandria, VA.

Dorman, K. S., Sinsheimer, J. S., and Kaplan, A. H. (1999). Estimating confidence in the inference of HIV recombination. In *ASA Proceedings of the Biometrics Section*, pp. 129–134. American Statistical Association, Alexandria, VA.

Draisma, G., de Haan, L., Peng, L., and Pereira, T. T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes* **2**, 367–404.

Ernst, M. D., and Hutson, A. D. (1999b). Exact bootstrap moments of an *L*-estimator. In *ASA Proceedings of the Statistical Computing Section*, pp. 123–126, American Statistical Association, Alexandria, VA.

Fachin, S., and Venanzoni, G. (1999). Patterns of internal migrations in Italy, 1970–1993. In *ASA Proceedings of the Business and Economic Statistics Section*, pp. 76–81. American Statistical Association, Alexandria, VA.

Famoye, F. (1999). EDF tests for the generalized Poisson distribution. *J. Statist. Comput. Simul.* **63**, 159–168.

Ferrari, S. L. P., and Cribari-Neto, F. (1999). On the robustness of analytical and bootstrap corrections to score tests in regression models. *J. Statist. Comput. Simul.* **64**, 177–191.

Feuerverger, A., Robinson, J., and Wong, A. (1999). On the relative accuracy of certain bootstrap procedures. *Can. J. Statist.* **27**, 225–236.

Fischer, I., and Harvey, N. (1999). Combining forecasts: What information do judges need to outperform the simple average? *Int. J. Forecast.* **15**, 227–246.

Genç, A. (1999). A simulation study of the bias of parameter estimators in multivariate nonlinear models. *Hacettepe Bull. Natural Sci. Eng., Ser. B: Math. Statist.* **28**, 105–114.

Godfrey, L. G., and Orme, C. D. (1999). The robustness, reliability and power of heteroskedasticity tests. *Econ. Rev.* **18**, 169–194.

Gombay, E., and Horváth, L. (1999). Change-points and bootstrap. *EnvironMetrics* **10**, 725–736.

Grigoletto, M. (1999). Bootstrap prediction intervals for autoregressive models fitted to non-autoregressive processes (Italian). In *Proceedings of the 1998 Conference of the Italian Statistical Society*, Vol. 2, pp. 841–848. *Societa Italiana di Statistica. Italian Statistical Society*, Rome.

Guillou, A. (1999a). Efficient weighted bootstraps for the mean. *J. Statist. Plann. Inf.* **77**, 11–35.

Guillou, A. (1999b). Weighted bootstraps for the variance. *J. Statist. Plann. Inf.* **81**, 113–120.

Guo, J.-H. (1999). A nonparametric test for the parallelism of two first-order autoregressive processes. *Aust. N. Z. J. Statist.* **41**, 59–65.

Halekoh, U., and Schweizer, K. (1999). Analysis of the stability of clusters of variables via bootstrap classification in the information age. In *Proceedings of the 22nd Annual GfKl Conference* (W. Gaul, and H. Locarek–Junge, editors), pp. 171–178. Springer-Verlag, Berlin.

Hall, P., Martin, M. A., and Sun, S. (1999). Monte Carlo approximation to Edgeworth expansions. *Can. J. Statist.* **27**, 579–584.

Hall, P., Peng, L., and Tajvidi, N. (1999). On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika* **86**, 871–880.

Hall, P., and Presnell, B. (1999a). Density estimation under constraints. *J. Comput. Graph. Statist.* **8**, 259–277.

Hall, P., and Presnell, B. (1999b). Intentionally biased bootstrap methods. *J. R. Statist. Soc. B* **61**, 143–158.

Hall, P., and Presnell, B. (1999c). Biased bootstrap methods for reducing the effects of contamination. *J. R. Statist. Soc. B* **61**, 661–680.

Hall, P., and Turlach, B. A. (1999). Reducing bias in curve estimation by use of weights. *Comput. Statist. Data Anal.* **30**, 67–86.

Hall, P., Wolff, R. C. L., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Am. Statist. Assoc.* **94**, 154–163.

Hazelton, M. L. (1999). An optimal local bandwidth selector for kernel density estimation. *J. Statist. Plann. Inf.* **77**, 37–50.

Hellmann, J. J., and Fowler, G. W. (1999). Bias, precision, and accuracy of four measures of species richness. *Ecol. Appl.* **9**, 824–834.

Hernández-Flores, C. N., Artiles-Romero, J., and Saavedra-Santana, P. (1999). Estimation of the population spectrum with replicated time series. *Comput. Statist. Data Anal.* **30**, 271–280.

Hesterberg, T. C. (1999). Bootstrap tilting confidence intervals and hypothesis tests. In *Computing Science and Statistics. Models, Predictions, and Computing*. Proceedings of the 31st Symposium on the Interface (K. Berk and Pourahmadi, Mohsen editors), pp. 389–393. Interface Foundation of North America, Fairfax Station, VA.

Hjorth, J. S. U. (1999). *Computer Intensive Statistical Methods: Validation Model Selection and Bootstrap*. Chapman & Hall, London.

Ichikawa, M., and Konishi, S. (1999). Model evaluation and information criteria in covariance structure analysis. *Brit. J. Math. Statist. Psych.* **52**, 285–302.

Iturria, S. J., Carroll, R. J., and Firth, D. (1999). Polynomial regression and estimating functions in the presence of multiplicative measurement error. *J. R. Statist. Soc. B* **61**, 547–561.

Jeong, K. M. (1999). Change-point estimation and bootstrap confidence regions in Weibull distribution. *J. Kor. Statist. Soc.* **28**, 359–370.

Johnson, P., and Cohen, F. (1999). Simple estimate and bootstrap confidence intervals about the estimated mean of the lytic unit at 20% cytotoxicity (percentile method, the BC_a and bootstrap-*t*). In Proceedings of the Twenty-Fourth Annual SAS Users Group International Conference, pp. 1688–1693. SAS Institute, Inc. Cary, NC.

Kakizawa, Y. (1999). Valid Edgeworth expansions of some estimators and bootstrap confidence intervals in first-order autoregression. *J. Time Series Anal.* **20**, 343–359.

Kaufman, S. (1999). Using the bootstrap to estimate the variance from a single systematic PPS sample. In *ASA Proceedings of the Section on Survey Research Methods*, pp. 683–688, American Statistical Association, Alexandria, VA.

Kazimi, C., and Brownstone, D. (1999). Bootstrap confidence bands for shrinkage estimators. *J. Econ.* **90**, 99–127.

Kilian, L. (1999). Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *J. Appl. Econ.* **14**, 491–510.

Kim, H. (1999). Variable selection in classification trees. In *ASA Proceedings of the Statistical Computing Section*, pp. 216–221. American Statistical Association, Alexandria, VA.

Kim, J. H. (1999). Asymptotic and bootstrap prediction regions for vector autoregression. *Int. J. Forecast.* **15**, 393–403.

Kim, T. Y. (1999). Block bootstrapped empirical process for dependent sequences. *J. Kor. Statist. Soc.* **28**, 253–264.

Kitagawa, G., and Konishi, S. (1999). Generalized information criterion (GIC) and the bootstrap (Japanese). *Proc. Inst. Statist. Math.* **47**, 375–394.

Klar, B. (1999). Goodness-of-fit tests for discrete models based on the integrated distribution function. *Metrika* **49**, 53–69.

Knautz, H. (1999). Nonlinear unbiased estimation in the linear regression model with nonnormal disturbances. *J. Statist. Plann. Inf.* **81**, 293–309.

Knight, K. (1999). Asymptotics for L_1-estimators of regression parameters under heteroscedasticity. *Can. J. Statist.* **27**, 497–507.

Kulperger, R. J. (1999). Countable state Markov process bootstrap. *J. Statist. Plann. Inf.* **76**, 19–29.

Kwon, K.-Y., and Huang, W.-M. (1999). Bootstrap on the GARCH models. In *ASA Proceedings of the Business and Economic Statistics Section*, pp. 257–262. American Statistical Association, Alexandria, VA.

Lahiri, S. N. (1999a). Theoretical comparison of block bootstrap methods. *Ann. Statist.* **27**, 386–404.

Lahiri, S. N. (1999b). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probab. Th. Rel. Fields* **114**, 55–84.

Lahiri, S. N. (1999c). Resampling methods for spatial prediction. *Computing Science and Statistics. Models, Predictions, and Computing. Proceedings of the 31st Symposium on the Interface* (K. Berk, and M. Pourahmadi editors), pp. 462–466. Interface Foundation of North America, Fairfax Station, VA.

Lahiri, S. N. (1999d). On second order properties of the stationary bootstrap method for studentized statistics. In *Asymptotics, Nonparametrics, and Time Series* (S. Ghosh editor), pp. 683–712, Marcel Dekker, New York.

Lahiri, S. N., Kaiser, M. S., Cressie, N., and Hsu, N. J. (1999). Prediction of spatial cumulative distribution functions using subsampling (with discussion). *J. Am. Statist. Assoc.* **94**, 86–110.

La Rocca, M. (1999). Conditional least squares and model-based bootstrap inference in bilinear models. *Quad. Statist.* **1**, 155–174.

Lee, D. S., and Chia, K. K. (1999). Recursive estimation for noise-free state models by Monte Carlo filtering. In *ASA Proceedings of the Section on Bayesian Statistical Science*, pp. 21–26. American Statistical Association, Alexandria, VA.

Lee, S. M. S. (1999). On a class of *m* out of *n* bootstrap confidence intervals. *J. R. Statist. Soc. B* **61**, 901–911.

Lee, S. M. S., and Young, G. A. (1999a). The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J. R. Statist. Soc. B* **61**, 353–366.

Lee, S. M. S., and Young, G. A. (1999b). Nonparametric likelihood ratio confidence intervals. *Biometrika* **86**, 107–118.

Leisch, F., and Hornik, K. (1999). Stabilization of *K*-means with bagged clustering. In *ASA Proceedings of the Statistical Computing Section*, pp. 174–179. American Statistical Association, Alexandria, VA.

Li, G., Tiwari, R. C., and Wells, M. T. (1999). Semiparametric inference for a quantile comparison function with applications to receiver operating characteristic curves. *Biometrika* **86**, 487–502.

Li, Q. (1999). Nonparametric testing the similarity of two unknown density functions: Local power and bootstrap analysis. *J. Nonparametric Statist.* **11**, 189–213.

Lillegård, M., and Engen, S. (1999). Exact confidence intervals generated by conditional parametric bootstrapping. *J. Appl. Statist.* **26**, 447–459.

Linden, M. (1999). Estimating effort function with semiparametric model. *Comput. Statist.* **14**, 501–513.

Luh, W.-M., and Guo, J.-H. (1999). A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under non-normality and inequality of variances. *Bri. J. Math. Statist. Psych.* **52**, 303–320.

Lumley, T., and Heagerty, P. (1999). Weighted empirical adaptive variance estimators for correlated data regression. *J. R. Statist. Soc. B* **61**, 459–477.

Minnotte, M. C. (1999). Nonparametric likelihood ratio testing of multimodality for discrete and binned data. In *ASA Proceedings of the Statistical Computing Section*, pp. 222–227. American Statistical Association, Alexandria, VA.

Munk, A., and Czado, C. (1999). A completely nonparametric approach to population bioequivalence in crossover trials. Preprint No.261, Preprint Series of the Faculty of Mathematics, Ruhr-Unversität Bochum.

Obenchain, R. L. (1999). Resampling and multiplicity in cost-effectiveness inference. *J. Biopharm. Statist.* **9**, 563–582.

Obenchain, R. L., and Johnstone, B. M. (1999). Mixed-model imputation of cost data for early discontinuers from a randomized clincial trial. *Drug Inf. J.* **33**, 191–209.

Oman, S. D., Meir, N., and Haim, N. (1999). Comparing two measures of creatinine clearance: An application of errors-in-variables and bootstrap techniques. *J. R. Statist. Soc. C* **48**, 39–52.

Pal, S. (1999). Performance evaluation of a bivariate normal process. *Quality Eng.* **11**, 379–386.

Pan, W. (1999). Bootstrapping likelihood for model selection with small samples. *J. Comput. Graph. Statist.* **8**, 687–698.

Paparoditis, E., and Politis, D. N. (1999). The local bootstrap for periodogram statistics. *J. Time Series Anal.* **20**, 193–222.

Park, D., and Willemain, T. R. (1999). The threshold bootstrap and threshold jackknife. *Comput. Statist. Data Anal.* **31**, 187–202.

Parke, J., Holford, N. H. G., and Charles, B. G. (1999). A procedure for generating bootstrap samples for the validation of nonlinear mixed-effects population models. *Comput. Methods Prog. Biomed.* **59**, 19–29.

Pettitt, A. N., and Choy, S. L. (1999). Bivariate binary data with missing values: Analysis of a field experiment to investigate chemical attractants of wild dogs. *J. Agr. Biol. Environ. Statist.* **4**, 57–76.

Pigeon, J. G., Bohidar, N. R., Zhang, Z., and Wiens, B. L. (1999). Statistical models for predicting the duration of vaccine-induced protection. *Drug Inf. J.* **33**, 811–819.

Polansky, A. M. (1999). Upper bounds on the true coverage of bootstrap percentile type confidence intervals. *Am. Statist.* **53**, 362–369.

Politis, D. N., Paparoditis, E., and Romano, J. P. (1999). Resampling marked point processes. In *Multivariate Analysis, Design of Experiments, and Survey Sampling: A Tribute to J. N. Srivastava* (S. Ghosh, editor). Marcel Dekker, New York.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*, Springer-Verlag, Berlin.*

Ratnaparkhi, M. V., and Waikar, V. B. (1999). Bootstrapping the two-stage shrinkage estimator of the mean of an exponential distribution. In *Computing Science and Statistics. Models, Predictions, and Computing. Proceedings of the 31st Symposium on the Interface* (K. Berk and M. Pourahmadi, editors), pp. 394–397. Interface Foundation of North America, Fairfax Station, VA.

Rius, R., Aluja-Banet, T., and Nonell, R. (1999). File grafting in market research. *Appl. Stoch. Models Bus. Ind.* **15**, 451–460.

Rothenberg, L., and Gullen, J. (1999). Assessing the appropriateness of different estimators for competing models with Likert scaled data using bootstrapping. In *ASA Proceedings of the Section on Government Statistics and Section on Social Statistics*, pp. 432–434. American Statistical Association, Alexandria, VA.

Sakov, A., and Bickel, P. J. (1999). Choosing m in the m out of n bootstrap. In *ASA Proceedings of the Section on Bayesian Statistical Science*, pp. 125–128. American Statistical Association, Alexandria, VA.*

Samawi, H. M., and Abu Awwad, R. K. (1999). Power estimation for two-sample tests using balanced resampling. *Commun. Statist. Th. Meth.* **28**, 1073–1092.

Sánchez-Sellero, C., González-Manteiga, W., and Cao, R. (1999). Bandwidth selection in density estimation with truncated and censored data. *Ann. Inst. Statist. Math.* **51**, 51–70.

Sauerbrei, W. (1999). The use of resampling methods to simplify regression models in medical statistics. *J. R. Statist. Soc. C, Appl. Statist.* **48**, 313–329.

Seaver, B., Triantis, K., and Hoopes, B. (1999). Complementary tools for influential subsets in regression: Fuzzy methods, regression diagnostics, and bootstrapping. In *ASA Proceedings of the Statistical Computing Section*, pp. 117–122. American Statistical Association, Alexandria, VA.

Simar, L., and Wilson, P. W. (1999). Estimating and bootstrapping Malmquist indices. *Eur. J. Op. Res.* **115**, 459–471.

Snethlage, M. (1999). Is bootstrap really helpful in point process statistics? *Metrika* **49**, 245–255.

Sødahl, N., and Godtliebsen, F. (1999). Semiparametric estimation of extremes. *Commun. Statist. Simul. Comput.* **28**, 569–595.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, New York.*

Streitberg, B. (1999). Exploring interactions in high-dimensional tables: A bootstrap alternative to log-linear models. *Ann. Statist.* **27**, 405–413.

Stute, W. (1999). Reply to comment on "Bootstrap approximations in model checks for regression." *J. Am. Statist. Assoc.* **94**, 660.

Tang, B. (1999). Balanced bootstrap in sample surveys and its relationship with balanced repeated replication. *J. Statist. Plann. Inf.* **81**, 121–127.

Tarpey, T. (1999). Self-consistency and principal component analysis. *J. Am. Statist. Assoc.* **94**, 456–467.

Tasker, G. D. (1999). Bootstrapping periodic ARMA model to forecast streamflow at multiple sites. In *Computing Science and Statistics. Models, Predictions, and Computing. Proceedings of the 31st Symposium on the Interface* (K. Berk and M. Pourahmadi, editors), pp. 296–299. Interface Foundation of North America, Fairfax Station, VA.

Taylor, C. C. (1999). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika* **76**, 705–712.

Tibshirani, R., and Knight, K. (1999a). Model search by bootstrap "bumping," *J. Comput. Graph. Statist.* **8**, 671–686.

Tibshirani, R., and Knight, K. (1999b). The covariance inflation criterion for adaptive model selection. *J. R. Statist. Soc. B* **61**, 529–546.

Timmer, J., Lauk, M., Vach, W., and Lücking, C. H. (1999). A test for a difference between spectral peak frequencies. *Comput. Statist. Data Anal.* **30**, 45–55.

Toktamis, Ö., Cula, S. G., and Kurt, S. (1999). Comparison of the bandwidth selection methods for kernel estimation of probability density function. *J. Turk. Statist. Assoc.* **2**, 107–121.

Tu, W., and Owen, W. J. (1999). A bootstrap procedure based on maximum likelihood summarization. In *ASA Proceedings of the Statistical Computing Section*, pp. 127–132. American Statistical Association, Alexandria, VA.

Turkheimer, F., Pettigrew, K., Sokoloff, L., and Schmidt, K. (1999). A minimum variance adaptive technique for parameter estimation and hypothesis testing. *Commun. Statist. Simul. Comput.* **28**, 931–956.

Vasdekis, V. G. S., and Trichopoulou, A. (1999). Bootstrap CI in estimated curves using penalized least squares; an example to household budget surveys data. In *Proceedings of the 12th Conference of the Greek Statistical Institute*, pp. 51–59. Greek Statistical Institute, Greece.

Vos, P. W., and Chenier, T. C. (1999). Trimmed conditional confidence intervals for a shift between two populations. *Commun. Statist. Simul. Comput.* **28**, 99–114.

Wang, J. (1999). Artificial likelihoods for general nonlinear regressions (Japanese). *Proc. Inst. Statist. Math.* **47**, 49–61.

Wassmer, G., Reitmeir, P., Kieser, M., and Lehmacher, W. (1999). Procedures for testing multiple endpoints in clinical trials: An overview. *J. Statist. Plann. Inf.* **82**, 69–81.

Wilcox, R. R. (1999). Testing hypotheses about regression parameters when the error term is heteroscedastic. *Biomed. J.* **41**, 411–426.

Wilcox, R. R. (1999). Comment on "Bootstrap approximations in model checks for regression." *J. Am. Statist. Assoc.* **94**, 659–660.

Wilcox, R. R., and Muska, J. (1999a). Measuring effect size: A non-parametric analogue of ^{2}. *Brit. J. Math. Statist. Psych.* **52**, 93–110.

Wilcox, R. R., and Muska, J. (1999b). Tests of hypotheses about regression parameters when using a robust estimator. *Commun. Statist. Th. Meth.* **28**, 2201–2212.

Winsberg, S., and De Soete, G. (1999). Latent class models for time series analysis. *Appl. Stoch. Models Bus. Ind.* **15**, 183–194.

Yeo, D., Mantel, H., and Liu, T.-P. (1999). Bootstrap variance estimation for the National Population Health Survey. In *ASA Proceedings of the Section on Survey Research Methods*, pp. 778–783, American Statistical Association, Alexandria, VA.

Zarepour, M., and Knight, K. (1999). Bootstrapping unstable first order autoregressive process with errors in the domain of attraction of stable law. *Commun. Statist. Stoch. Models* **15**, 11–27.

Zhang, B. (1999). Bootstrapping with auxiliary information. *Canad. J. Statist.* **27**, 237–249.

2000

Abdelhafez, M. E. M., and Ismail, M. A. (2000). Interval estimation for amplitude-dependent exponential autoregressive (EXPAR) models. *Egyptian Statist. J.* **44**, 1–10.

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Soc. Meth. Res.* **28**, 301–309.

Almudevar, A., Bhattacharya, R. N., and Sastri, C. C. A. (2000). Estimating the probability mass of unobserved support in random sampling. *J. Statist. Plann. Inf.* **91**, 91–105.

Alonso, A. M., Peña, D., and Romo, J. (2000). Sieve bootstrap prediction intervals. In *COMPSTAT—Proceedings in Computational Statistics, 14th Symposium* (J. G. Bethlehem and P. G. M. van der Heijden, editors). pp. 181–186. Physica-Verlag Ges.m.b.H Heidelberg.

Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68**, 399–405.

Andrews, D. W. K., and Buchinsky, M. (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica* **68**, 23–51.

Andronov, A., and Merkuryev, Y. (2000). Optimization of statistical sample sizes in simulation. *J. Statist. Plann. Inf.* **85**, 93–102.

Antoch, J., and Hušková, M. (2000). Bayesian-type estimators of change points. *J. Statist. Plann. Inf.* **91**, 195–208.

Azzalini, A., and Hall, P. (2000). Reducing variability using bootstrap methods with qualitative constraints. *Biometrika* **87**, 895–906.

Babu, G. J., Pathak, P. K., and Rao, C. R. (2000). Consistency and accuracy of the sequential bootstrap. In *Statistics for the 21st Century: Methodologies for Application to the Future* (C. R. Rao and G. Szekely, editors), pp. 21–31, Marcel Dekker, New York.

Baker, S. G. (2000). Identifying combinations of cancer markers for further study as triggers of early intervention. *Biometrics* **56**, 1082–1087.

Barabesi, L. (2000). On the exact bootstrap distribution of linear statistics. *Metron* **58**, 53–63.

Barber, J. A., and Thompson, S. G. (2000). Analysis of cost data in randomized trials: An application of the nonparametric bootstrap. *Statist. Med.* **19**, 3219–3236.

Bartels, K. (2000). A linear approximation to the wild bootstrap in specification testing. In *COMPSTAT—Proceedings in Computational Statistics, 14th Symposium* (J. G. Bethlehem and P. G. M. van der Heijden, editors). pp. 205–210. Physica-Verlag Ges.m.b.H, Heidelberg.

Bartels, K., Boztug, Y., and Müller, M. (2000). Testing the multinomial logit model. In *Classification and Information Processing at the Turn of the Millennium* (R. Decker and W. Gaul, editors). pp. 296–303. Springer-Verlag Inc. Berlin.

Belyaev, Y., and Sjöstedt-de Luna, S. (2000). Weakly approaching sequences of random distributions. *J. Appl. Probab.* **37**, 807–822.

Benkwitz, A., Lütkepohl, H., and Neumann, M. H. (2000). Problems related to confidence intervals for impulse responses of autoregressive processes. *Econ. Rev.* **19**, 69–103.

Bergkvist, E., and Johansson, P. (2000). Weighted derivative estimation of quantal response models: Simulations and applications to choice of truck freight carrier. *Comput. Statist.* **15**, 485–510.

Berkovits, I., Hancock, G. R., and Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Edu. Psych. Meas.* **60**, 877–892.

Berkowitz, J., and Kilian, L. (2000). Recent developments in bootstrapping time series. *Econ. Rev.* **19**, 1–48.

Berry, V., Gascuel, O., and Caraux, G. (2000). Choosing the tree which actually best explains the data: Another look at the bootstrap in phylogenetic reconstruction. *Comput. Statist. Data Anal.* **32**, 273–283.

Bertail, P., Politis, D. N., and Rhomari, N. (2000). Subsampling continuous parameter random fields and a Bernstein inequality. *Statistics* **33**, 367–392.

Bickel, P. J., and Ritov, Y. (2000). Non- and semiparametric statistics: Compared and contrasted. *J. Statist. Plann. Inf.* **91**, 209–228.

Bilder, C. R., Loughin, T. M., and Nettleton, D. (2000). Multiple marginal independence testing for pick any *c* variables. *Commun. Statist. Simul. Comput.* **29**, 1285–1316.

Bittanti, S., and Lovera, M. (2000). Bootstrap-based estimates of uncertainty in subspace identification methods. *Automatica* **36**, 1605–1615.

Boos, D. D., and Zhang, J. (2000). Monte Carlo evaluation of resampling-based hypothesis tests. *J. Am. Statist. Assoc.* **95**, 486–492.

Borkowf, C. B. (2000). A new nonparametric method for variance estimation and confidence interval construction for Spearman's rank correlation. *Comput. Statist. Data Anal.* **34**, 219–241.

Bouza, C. N. (2000). Adaptive sampling under random contacts. *Rev. Invest. Op. (Havana)* **21**, 38–45.

Bowman, A. W., and Wright, E. M. (2000). Graphical exploration of covariate effects on survival data through nonparametric quantile curves. *Biometrics* **56**, 563–570.

Broström, G., and Nilsson, L. (2000). Acceptance-rejection sampling from the conditional distribution of independent discrete random variables, given their sum. *Statistics* **34**, 247–257.

Brumback, B. A., Ryan, L. M., Schwartz, J. D., Neas, L. M., Stark, P. C., and Burge, H. A. (2000). Transitional regression models, with application to environmental time series. *J. Am. Statist. Assoc.* **95**, 16–27.

Buckland, S. T., Augustin, N. H., Trenkel, V. M., Elston, D. A., and Borchers, D. L. (2000). Simulated inference, with applications to wildlife population assessment. *Metron* **58**, 3–22.

Bühlmann, P. (2000). Model selection for variable length Markov chains and tuning the context algorithm. *Ann. Inst. Statist. Math.* **52**, 287–315.

Burke, M. D. (2000). Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. *Statist. Probab. Lett.* **46**, 13–20.

Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Am. Statist. Assoc.* **95**, 888–902.

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Am. Statist. Assoc.* **95**, 941–956.

Canty, A. J., and Davison, A. C. (2000). Comment on "The estimating function bootstrap." *Can. J. Statist.* **28**, 489–493.

Carpenter, J., and Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statist. Med.* **19**, 1141–1164.

Carriere, J. F. (2000). Non-parametric confidence intervals of instantaneous forward rates. *Inst. Math. Econ.* **26**, 193–202.

Chatterjee, S., and Bose, A. (2000). Variance estimation in high-dimensional regression models. *Statist. Sin.* **10**, 497–515.

Chao, A., Chu, W., and Hsu, C.-H. (2000). Capture-recapture when time and behavioral response affect capture probabilities. *Biometrics* **56**, 427–433.

Chen, J., Rao, J. N. K., and Sitter, R. R. (2000). Efficient random imputation for missing data in complex surveys. *Statist. Sin.* **10**, 1153–1169.

Chen, J. J., Lin, K. K., Huque, M., and Arani, R. B. (2000). Weighted p-value adjustments for animal carcinogenicity trend test. *Biometrics* **56**, 586–592.

Chen, M., and Kianifard, F. (2000). A nonparametric procedure associated with a clinically meaningful efficacy measure. *Biostatistics* **1**, 293–298.

Chen-Mok, M., and Sen, P. K. (2000). Nondifferentiable errors in beta-compliance integrated logistic models: Numerical results. *Commun. Statist. Simul. Comput.* **29**, 1149–1164.

Choi, D., and Yoon, J. (2000). On the classification by an improved pairwise coupling algorithm. *Kor. J. Appl. Statist.* **13**, 415–425.

Choi, E., and Hall, P. (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *J. R. Statist. Soc. B* **62**, 461–477.

Choi, E., Hall, P., and Presnell, B. (2000). Rendering parametric procedures more robust by empirically tilting the model. *Biometrika* **87**, 453–465.

Christman, M. C., and Pontius, J. S. (2000). Bootstrap confidence intervals for adaptive cluster sampling. *Biometrics* **56**, 503–510.

Chuang, C.-S., and Lai, T. L. (2000). Hybrid resampling methods for confidence intervals. *Statist. Sin.* **10**, 1–50.

Chung, H.-C., and Han, C.-P. (2000). Discriminant analysis when a block of observations is missing. *Ann. Inst. Statist. Math.* **52**, 544–556.

Claeskens, G., and Aerts, M. (2000). Bootstrapping local polynomial estimators in likelihood-based models. *J. Statist. Plann. Inf.* **86**, 63–80.

Clemons, T. E., andBradley, E. L., Jr (2000). A nonparametric measure of the overlapping coefficient. *Comput. Statist. Data Anal.* **34**, 51–61.

Colosimo, E. A., Silva, A. F., and Cruz, F. R. B. (2000). Bias evaluation in the proportional hazards model. *J. Statist. Comput. Simul.* **65**, 191–201.

Concordet, D., and Nunez, O. G. (2000). Calibration for nonlinear mixed effects models: An application to the withdrawal time prediction. *Biometrics* **56**, 1040–1046.

Contreras, M., and Ryan, L. M. (2000). Fitting nonlinear and constrained generalized estimating equations with optimization software. *Biometrics* **56**, 1268–1271.

Cook, J. R., and Heyse, J. F. (2000). Use of an angular transformation for ratio estimation in cost-effectiveness analysis. *Statist. Med.* **19**, 2855–2866.

Cribari-Neto, F., and Zarkos, S. G. (2000). The size and power of analytical and bootstrap corrections to score test in regression models. *Commun. Statist. Theory Meth.* **29**, 279–289.

Csörgö, M., Horváth, L., and Kokoszka, P. (2000). Approximation for bootstrapped empirical processes. *Proc. Am. Math. Soc.* **128**, 2457–2464.

Csörgo, S., and Wu, W. B. (2000). Random graphs and the strong convergence of bootstrap means. *Comb. Probab. Comput.* **9**, 315–347.

Cuevas, A., Febrero, M., and Fraiman, R. (2000). Estimating the number of clusters. *Canad. J. Statist.* **28**, 367–382.

Cula, S. G., and Toktamis, Ö. (2000). Estimation of multivariate probability density function with kernel functions. *J. Turk. Statist. Assoc.* **3**, 29–39.

Dahlberg, M., and Johansson, E. (2000). An examination of the dynamic behaviour of local governments using GMM bootstrapping methods. *J. Appl. Econ.* **15**, 401–416.

Darken, P. F., Holtzman, G. I., Smith, E. P., and Zipper, C. E. (2000). Detecting changes in trends in water quality using modified Kendall's tau. *EnvironMetrics* **11**, 423–434.

Davidson, R. (2000). Comment on "Recent developments in bootstrapping time series." *Econ. Rev.* **19**, 49–54.

Davidson, R., and MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps? *Econ. Rev.* **19**, 55–68.

Davison, A. C., and Louzado-neto, F. (2000). Inference for the poly-Weibull model. *J. R. Statist. Soc. D, The Statistician* **49**, 189–196.

Davison, A. C., and Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. *J. R. Statist. Soc. B* **62**, 191–208.

De Martini, D. (2000). Smoothed bootstrap consistency through the convergence in Mallows metric of smooth estimates. *J. Nonparametric Statist.* **12**, 819–835.

Denham, M. C. (2000). Choosing the number of factors in partial least squares regression: Estimating and minimizing the mean squared error of prediction. *J. Chemometrics* **14**, 351–361.

Del Barrio, E., and Matrán, C. (2000). The weighted bootstrap mean for heavy-tailed distributions. *J. Th. Probab.* **13**, 547–569.

de Uña-Álvarez, J., González-Manteiga, W., and Cadarso-Suárez, C. (2000). Kernel distribution function estimation under the Koziol-Green model. *J. Statist. Plann. Inf.* **87**, 199–219.

DiCiccio, T. J., and Tibshirani, R. J. (2000). Comment on "The estimating function bootstrap." *Can. J. Statist.* **28**, 485–487.

Efron, B. (2000). The bootstrap and modern statistics. *J. Am. Statist. Assoc.* **95**, 1293–1296.

El Barmi, H., and Nelson, P. I. (2000). Three monotone density estimators from selection biased samples. *J. Statist. Comput. Simul.* **67**, 203–217.

El-Bassiouni, M. Y., and Abdelhafez, M. E. M. (2000). Interval estimation of the mean in a two-stage nested model. *J. Statist. Comput. Simul.* **67**, 333–350.

El-Nouty, C., and Guillou, A. (2000a). On the smoothed bootstrap. *J. Stat Plann. Inf.* **83**, 203–220.

El-Nouty, C., and Guillou, A. (2000b). On the bootstrap accuracy of the Pareto index. *Statist. Dec.* **18**, 275–289.

Emir, B., Wieand, S., Jung, S.-H., and Ying, Z. (2000). Comparison of diagnostic markers with repeated measurements. A non-parametric ROC curve approach. *Statist. Med.* **19**, 511–523.

Escolono, S., Golmard, J.-L., Korinek, A.-M., and Mallet, A. (2000). A multi-state model for evolution of intensive care unit patients: Prediction of nosocomial infections and death. *Statist. Med.* **19**, 3465–3482.

Fachin, S. (2000). Bootstrap and asymptotic tests of long-run relationships in cointegrated systems. Oxford Bull. *Econ. Statist.* **62**, 543–551.

Famoye, F. (2000). Goodness-of-fit tests for generalized logarithmic series distribution. *Comput. Statist. Data Anal.* **33**, 59–67.

Fan, J., Hsu, L., and Prentice, R. L. (2000). Dependence estimation over a finite bivariate failure time region. *Lifetime Data Anal.* **6**, 343–355.

Feiveson, A. H., and Kulkarni, P. M. (2000). Reliability of space-shuttle pressure vessels woth random batch effects. *Technometrics* **42**, 332–344.

Fiebig, D. G., and Kim, J. H. (2000). Estimation and inference in SUR models when the number of equations is large. *Econ. Rev.* **19**, 105–130.

Fricker, R. D., Jr., and Goodhart, C. A. (2000). Applying a bootstrap approach for setting reorder points in military supply systems. *Nav. Res. Logist.* **47**, 459–478.

Fuh, C. D., Fan, T. H., and Hung, W. L. (2000). Balanced importance resampling for Markov chains. *J. Statist. Plann. Inf.* **83**, 221–241.

Gail, M. H., Pfeiffer, R., van Houwelingen, H. C., and Carroll, R. J. (2000). On meta-analytic assessment of surrogate outcomes. *Biostatistics* **1**, 231–246.

Geskas, R. B. (2000). On the inclusion of prevalent cases in HIV/AIDS natural history studies through a marker-based estimate of time since sero conversion. *Statist. Med.* **19**, 1753–1764.

Ghattas, B. (2000). Aggregation of classification trees. *Rev. Statist. Appl.* **48**, 85–98.

Ghosh, S., and Beran, J. (2000). Two-sample T_3 plot: A graphical comparison of two distributions. *J. Comput. Graph. Statist.* **9**, 167–179.

Ghoudi, K., and McDonald, D. (2000). Cramér-von Mises regression. *Can. J. Statist.* **28**, 689–714.

Gigli, A., and Verdecchia, A. (2000). Uncertainty of AIDS incubatuib time and its effects on back-calculation estimates. *Statist. Med.* **19**, 1–11.

Gilbert, P. (2000). Developing an AIDS vaccine by sieving. *Chance* **13**, 16–21.

Godfrey, L. G., and Orme, C. D. (2000). Controlling the significance levels of prediction error tests for linear regression models. *Econ. J. Online* **3**, 66–83.

Godfrey, L. G., and Veall, M. R. (2000). Alternative approaches to testing by variable addition. *Econ. Rev.* **19**, 241–261.

Good, P. (2000). *Permutation Tests*, second edition. Springer-Verlag, New York.*

Grenier, M., and Léger, C. (2000). Bootstrapping regression models with BLUS residuals. *Can. J. Statist.* **28**, 31–43.

Grübel, R., and Pitts, S. M. (2000). Statistical aspects of perpetuities. *J. Mult. Anal.* **75**, 143–162.

Guillou, A. (2000). Bootstrap confidence intervals for the Pareto index. *Commun. Statist. Th. Meth.* **29**, 211–226.

Gürtler, N., and Henze, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution. *J. Statist. Plann. Inf.* **90**, 207–225.

Hafner, C. M., and Herwartz, H. (2000). Testing for linear autoregressive dynamics under heteroskedasticity. *Econ. J. Online* **3**, 177–197.

Hall, P., Härdle, W., Kleinow, T., and Schmidt, P. (2000). Semiparametric bootstrap approach to hypothesis tests and confidence intervals for the Hurst coefficient. *Statist. Inf. Stoch. Proc.* **3**, 263–276.

Hall, P., and Heckman, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28**, 20–39.

Hall, P., and Maesono, Y. (2000). A weighted bootstrap approach to bootstrap iteration. *J. R. Statist. Soc. B* **62**, 137–144.

Hall, P., Presnell, B., and Turlach, B. A. (2000). Reducing bias without prejudicing sign. *Ann. Inst., Statist. Math.* **52**, 507–518.

Han, J. H., Cho, J. J., and Leem, C. S. (2000). Bootstrap confidence limits for Wright's C_s Commun. *Statist. Th. Meth.* **29**, 485–505.

Hansen, B. E. (2000). Testing for structural change in conditional models. *J. Econ.* **97**, 93–115.

Hauptmann, M., Wellmann, J., Lubin, J. H., Rosenberg, P. S., and Kreienbrock, L., (2000). Analysis of exposure-time–response relationships using a spline weight function. *Biometrics* **56**, 1105–1108.

Hauschke, D., and Steinijans, V. W. (2000). The U. S. draft guidance regarding population and individual bioequivalence approaches: Comment by a research-based pharmaceutical company. *Statist. Med.* **19**, 2769–2774.

Heagerty, P. J., and Lumley, T. (2000). Window subsampling of estimating functions with application to regression models. *J. Am. Statist. Assoc.* **95**, 197–211.

Helmers, R. (2000). Miscellanea, inference on rare errors using asymptotic expansions and bootstrap calibration. *Biometrika* **87**, 689–694.

Herwartz, H. (2000). Weekday dependence of German stock market returns. *Appl. Stoch. Models Bus. Ind.* **16**, 47–71.

Horowitz, J. L., and Savin, N. E. (2000). Empirically relevant critical values for hypothesis tests: A bootstrap approach. *J. Econ.* **95**, 375–389.

Horváth, L., Kokoszka, P., and Steinebach, J. (2000). Approximations for weighted bootstrap processes with an application. *Statist. Probab. Lett.* **48**, 59–70.

Hu, F., and Hu, J. (2000). A note on breakdown theory for bootstrap methods. *Statist. Probab. Lett.* **50**, 49–53.

Hu, F., and Kalbfleisch, J. D. (2000a). The estimating function bootstrap. *Can. J. Statist.* **28**, 449–481.

Hu, F., and Kalbfleisch, J. D. (2000b). Reply to comments on "The estimating function bootstrap." *Can. J. Statist.* **28**, 496–499.

Huang, Y., Kirby, S. P. J., Harris, P., and Dearden, J. C. (2000). Interval estimation of the 90% effective dose: A comparison of bootstrap resampling methods with some large-sample approaches. *J. Appl. Statist.* **27**, 63–73.

Hung, W.-L. (2000). Bayesian bootstrap clones for censored Markov chains. *Biomed. J.* **42**, 501–510.

Hutchison, D., Morrison, J., and Felgate, R. (2000). Bootstrapping the effect of measurement errors in variables on regression results. In *ASA Proceedings of the Section on Survey Research Methods*, pp. 569–573. American Statistical Association, Alexandria, VA.

Hutson, A. D. (2000a). A composite quantile function estimator with applications in bootstrapping. *J. Appl. Statist.* **27**, 567–577.

Hutson, A. D. (2000b). Estimating the covariance of bivariate order statistics with applications. *Statist. Probab. Lett.* **48**, 195–203.

Hutson, A. D., and Ernst, M. D. (2000). The exact bootstrap mean and variance of an *L*-estimator *J. R. Statist. Soc. B* **62**, 89–94.

Jalaluddin, M., and Kosorok, M. R. (2000). Regional confidence bamds for ROC curves. *Statist. Med.* **19**, 493–509.

Jeng, S.-L., and Meeker, W. Q. (2000). Comparisons of approximate confidence intereval procedures for type I censored data. *Technometrics* **42**, 332–344.

Jhun, M., and Jeong, H.-C. (2000). Applications of bootstrap methods for categorical data analysis. *Comput. Statist. Data Anal.* **35**, 83–91.

Jiang, G., Wu, J., and Williams, G. R. (2000a). Bootstrap and likelihood confidence intervals for the incremental cost-effectiveness ratio. In *ASA Proceedings of the Biopharmaceutical Section*, pp. 105–110. American Statistical Association, Alexandria, VA.

Jiang, G., Wu, J., and Williams, G. R. (2000b). Fieller's interval and the bootstrap-Fieller interval for the incremental cost-effectiveness ratio. *Health Serv. Out. Res. Meth.* **1**, 291–303.

Jones, M. C. (2000). Rough-and-ready assessment of the degree and importance of smoothing in functional estimation. *Statist. Neerl.* **54**, 37–46.

Karian, Z. A., and Dudewicz, E. J. (2000). *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap methods*. Chapman & Hall, London.

Karlis, D., and Xekalaki, E. (2000). A simulation comparison of several procedures for testing the Poisson assumption. *J. R. Statist. Soc. D, The Statistician* **49**, 355–382.

Karlsson, S., and Löthgren, M. (2000). Computationally efficient double bootstrap variance estimation. *Comput. Statist. Data Anal.* **33**, 237–247.

Kaufman, S. (2000). Using the bootstrap to estimate the variance in a very complex sample design. In *ASA Proceedings of the Section on Survey Research Methods*, pp. 180–185. American Statistical Association, Alexandria, VA.

Kedem, B., and Kozintsev, B. (2000). Graphical bootstrap. In *ASA Proceedings of the Section on Statistics and the Environment*, pp. 30–32. American Statistical Association, Alexandria, VA.

Keselman, H. J., Kowalchuk, R. K., Algina, J., Lix, L. M., and Wilcox, R. R. (2000). Testing treatment effects in repeated measures designs: Trimmed means and bootstrapping. *Brit. J. Math. Statist. Psych.* **53**, 175–191.

Kester, A. D. M., and Buntinx, F. (2000). Meta-analysis of ROC curves. *Med. Dec. Making* **20**, 430–439.

Kilian, L., and Demiroglu, U. (2000). Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence. *J. Bus. Econ. Statist.* **18**, 40–50.

Kimanani, E. K., Lavigne, J., and Potvin, D. (2000). Numerical methods for the evaluation of individual bioequivalence approaches: Comments by a research-based pharmaceutical company. *Statist. Med.* **19**, 2775–2795.

Kundu, D., and Basu, S. (2000). Analysis of incomplete data in presence of competing risks. *J. Statist. Plann. Inf.* **87**, 221–239.

Kuonen, D. (2000). A saddlepoint approximation for the collector's problem. *Am. Statist.* **54**, 165–169.

Lee, D., and Priebe, C. (2000). Exact mean and mean squared error of the smoothed bootstrap mean integrated squared error estimatr. *Comput. Statist.* **15**, 169–181.

Lee, S. M. S. (2000a). Comment on "The estimating function bootstrap." *Can. J. Statist.* **28**, 494–495.

Lee, S. M. S. (2000a). Nonparametric confidence interval based on extreme bootstrap percentiles. *Statist. Sin.* **10**, 475–496.

Léger, C. (2000). Comment on "The estimating function bootstrap." *Can. J. Statist.* **28**, 487–489.

Li, H. (2000). The power of bootstrap based tests for parameters in cointegrating regressions. *Statist. Papers* **41**, 197–210.

Li, H., and Xiao, Z. (2000). On bootstrapping regressions with unit root processes. *Statist. Probab. Lett.* **48**, 261–267.

Li, X., Zhong, C., and Jing, B. (2000). Test for new better than used in convex ordering. *Commun. Statist. Th. Meth.* **29**, 2751–2760.

Liang, H., Härdle, W., and Sommerfeld, V. (2000). Booststrap approximation in a partially linear regression model. *J. Statist. Plann. Inf.* **91**, 413–426.

Liora, J., and Delgado-Rodriguez, M. (2000). A comparison of several procedures to estimate the confidence interval for attributable risk in case-control studies. *Statist. Med.* **19**, 1089–1099.

Liu, J.-P., and Ma, M.-C. (2000). On difference factor in assessment of dissolution similarity. *Commun. Statist. Th. Meth.* **29**, 1089–1113.

Luh, W.-M., and Guo, J.-H. (2000). Johnson's transformation two-sample trimmed t and its bootstrap method for heterogeneity and non-normality. *J. Appl. Statist.* **27**, 965–973.

Lütkepohl, H. (2000). Bootstrapping impulse responses in VAR analyses. In *COMPSTAT—Proceedings in Computational Statistics, 14th Symposium* (J. G. Bethehem and P. O. M. van der Heijden, editors), pp. 349–354. Physica-Verlag, Heidelberg.

Maasoumi, E., and Heshmati, A. (2000). Stochastic dominance amongst Swedish income distributions. *Econ. Rev.* **19**, 287–320.

Maharaj, E. A. (2000). Comparison of stationary time series using distribution-free methods. In *COMPSTAT—Proceedings in Computational Statistics, 14th Symposium* (J. G. Bethlehem and P. G. M. van der Heijden, editors), pp. 349–354. Physica-Verlag Ges.m.b.H, Heidelberg.

Mahoud, M., Mokhlis, N. A., and Ibrahim, S. A. N. (2000). Assessing the error in bootstrap estimates with dependent data. *Test* **9**, 471–486.

Mammen, E. (2000). Resampling methods for nonparametric regression. In *Smoothing and Regression: Approaches, Computation, and Application* (M. G. Schimek, editor), pp. 425–450. Wiley, New York.

Maugher, D. T., and Chinchilli, V. M. (2000). An alternative index for assessing profile similarity in bioequivalence trials. *Statist. Med.* **19**, 2855–2866.

McCullagh, P. (2000). Resampling and exchangeable arrays. *Bernoulli* **6**, 285–301.

McKnight, S. D., McKean, J. W., and Huitema, B. E. (2000). A double bootstrap method to analyze linear models with autoregressive error terms. *Psych. Meth.* **5**, 87–101.

McGee, D. L. (2000). Analyzing and synthesizing information from a multiple-study database. In *COMPSTAT—Proceedings in Computational Statistics, 14th Symposium* (J. G. Bethlehem and P. G. M. van der Heijden, editors), pp. 469–474. Physica-Verlag Ges.m.b.H, Heidelberg.

Moreno, M., and Romo, J. (2000). Bootstrap tests for unit roots based on LAD estimation. *J. Statist. Plann. Inf.* **83**, 347–367.

Mulekar, M. S., and Mishra, S. N. (2000). Confidence interval estimation of overlap: Equal means case. *Comput. Statist. Data Anal.* **34**, 121–137.

Muliere, P., and Walker, S. (2000). Neutral to the right processes from a predictive perspective: A review and new developments. *Metron* **58**, 13–30.

Neath, A. A., and Cavanaugh, J. E. (2000). A regression model selection criterion based on bootstrap bumping for use with resistant fitting. *Comput. Statist. Data Anal.* **35**, 155–169.

Nelson, P. I., and Kemp, K. E. (2000). Small sample set estimation of a baseline ranking. *Commun. Statist. Th. Meth.* **29**, 19–43.

Neumann, M. H., and Paparoditis, E. (2000). On bootstrapping L_2-type statistics in density testing *Statist. Probab. Lett.* **50**, 137–147.

Nguyen, V. T. (2000). On weak convergence of the bootstrap empirical process woth random resample size. *Vietnam J. Math.* **28**, 133–158.

Omar, R. Z., and Thompson, S. G. (2000). Analysis of a cluster randomized trial with binary outcome data using a multi-level model. *Statist. Med.* **19**, 2675–2688.

Pallini, A. (2000). Efficient bootstrap estimation of distribution functions. *Metron* **58**, 81–95.

Palmitesta, P., Provasi, C., and Spera, C. (2000). Confidence interval estimation for inequality indices of the Gini family. *Comput. Econ.* **16**, 137–147.

Pan, W. (2000a). A two-sample test with interval censored data via multiple imputation. *Statist. Med.* **19**, 1–11.

Pan, W. (2000b). Smooth estimation of the survival function for interval consored data. *Statist. Med.* **19**, 2611–2624.

Paparoditis, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scand. J. Statist.* **27**, 143–176.

Paparoditis, E., and Politis, D. N. (2000a). The local bootstrap for kernel estimators under general dependence conditions. *Ann. Inst. Statist. Math.* **52**, 139–159.

Paparoditis, E., and Politis, D. N. (2000b). Large-sample inference on the general AR(1) model. *Test* **9**, 487–509.

Park, H.-I, and Na, J.-H. (2000). Bootstrap median tests for right censored data. *J. Kor. Statist. Soc.* **29**, 423–433.

Pawitan, Y. (2000). Computing empirical likelihood from the bootstrap. *Statist. Probab. Lett.* **47**, 337–345.

Percival, D. B., Sardy, S., and Davison, A. C. (2000). Wave strapping time series: Adaptive wavelet based bootstrapping. In *Nonlinear and Nonstationary Signal Processing (Cambridge 1998)*, pp. 442–471. Cambridge University Press, Cambridge.

Phipps, M. C. (2000). Power surfaces. *Math. Scientist* **25**, 100–104.

Plstt, R. W., Hanley, J. A., and Yang, H. (2000). Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test. *Statist. Med.* **19**, 313–322.

Polansky, A. M. (2000). Stabilizing bootstrap-*t* confidence intervals for small samples. *Can. J. Statist.* **28**, 501–516.

Politis, K., and Pitts, S. M. (2000). Nonparametric estimation in renewal theory. II: Solutions of renewal-type equations. *Ann. Statist.* **28**, 88–115.

Psaradakis, Z. (2000). Bootstrap tests for unit roots in seasonal autoregressive models. *Statist. Probab. Lett.* **50**, 389–395.

Rao, J. S. (2000). Bootstrapping to assess and improve atmospheric prediction models. *Data Mining and Knowledge Discovery* **4**, 29–41.

Ren, Z., and Chen, M. (2000). The Edgeworth expansion and the Bootstrap approximation for an *L*-statistic. *Chin. J. Appl. Probab. Statist.* **16**, 113–124.

Robins, J. M., Ventura, V., and van der Vaart, A. (2000). Asymptotic distribution of *p*-Values in composite null models. *J. Am. Statist. Assoc.* **95**, 1143–1156.

Romano, J. P., and Wolf, M. (2000). A more general central limit theorem for *m*-dependent random variables with unbounded *m*. *Statist. Probab. Lett.* **47**, 115–124.

Sakov, A., and Bickel, P. J. (2000). An Edgeworth expansion for the *m* out of *n* bootstrapped median. *Statist. Probab. Lett.* **49**, 217–223.

Scagni, A. (2000). Bootstrap goodness-of-fit tests for complex survey samples. In *Data Analysis, Classification, and Related Methods* (H. A. L. Kiers, J.-P. Rasson, P. J. F. Groenen, and M. Schader, editors). Springer-Verlag Inc. Berlin.

Schiavo, R. A., and Hand, D. J. (2000). Ten more years of error rate research. *Int. Statist. Rev.* **68**, 295–310.

Semenciw, R. M., Le, N. D., Marrett, L. D., Robson, D. L., Turner, D., and Walter, S. D. (2000). Methodological issues in the development of the Canadian Cancer Incidence Atlas. *Statist. Med.* **19**, 2437–2499.

Shao, Q. (2000). Estimation for hazardous concentrations based on NOEC toxicity data: An alternative approach. *EnvironMetrics* **11**, 583–595.

Shao, J., Chow, S.-C., and Wang, B. (2000). The bootstrap procedure for individual bioequivalence. *Statist. Med.* **19**, 2741–2754.

Shao, J., Kübler, J., and Pigeot, I. (2000). Consistency of the bootstrap procedure in individual bioequivalence. *Biometrika* **87**, 573–585.*

Shen, P.-S. (2000). Confidence interval estimation of mean survival time. *J. Ch. Statist. Assoc.* **38**, 73–82.

Simar, L., and Wilson, P. W. (2000). A general methodology for bootstrapping in nonparametric frontier models. *J. Appl. Statist.* **27**, 779–802.

Simonetti, N. (2000). Variance estimation for orthogonal series estimators of probability densities. *Metron* **58**, 111–120.

Sjöstedt, S. (2000). Resampling *m*-dependent random variables with applications to forecasting. *Scand. J. Statist.* **27**, 543–561.

Small, C. G., Wang, J., and Yang, Z. (2000). Eliminating multiple root problems in estimation. *Statist. Sci.* **15**, 313–332.

Solow, A. R. (2000). Comment on "Multiple comparisons of entropies with application to dinosaur biodiversity." *Biometrics* **56**, 1272–1273.

Steele, B. M., and Patterson, D. A. (2000). Ideal bootstrap estimation of expected prediction error for *k*-nearest neighbor classifiers: Applications for classification error assessment. *Statists. Comput.* **10**, 349–355.

Stein, M. L., Quashnock, J. M., and Loh, J. M. (2000). Estimating the *K* function of a point process with an application to cosmology. *Ann. Statist.* **28**, 1503–1532.

Stute, W., González-Manteiga, W., and Sánchez, S. C. (2000). Nonparametric model checks in censored regression. *Commun. Statist. Th. Meth.* **29**, 1611–1629.

Tamura, R. N., Faries, D. E., and Feng, J. (2000). Comparing time to onset of response in antidepressant clinical trials using the cure model and the Cramer–Von Mises test. *Statist. Med.* **19**, 2169–2184.

Tan, W.-Y., and Ye, Z. (2000). Some state space models of HIV epidemic and its applications for the estimation of HIV infection and incubation. *Commun. Statist. Th. Meth.* **29**, 1059–1088.

Thomas, G. E. (2000). Use of the bootstrap in robust estimation of location. *J. R. Statist. Soc. D. The Statistician* **49**, 63–77.

Thorpe, D. P., and Holland, B. (2000). Some multiple comparison procedures for variances from non-normal populations. *Comput. Statist. Data Anal.* **35**, 171–199.

Tsujitani, M., and Koshimizu, T. (2000). Bootstrapping neural discriminant model. In *COMPSTAT—Proceedings in Computational Statistics, 14th Symposium* (J. G. Bethlehem and P. G. M. van der Heijden, editors), pp. 475–480. Physica-Verlag Ges.m.b.H, Heidelberg.

Tu, W., and Zhou, X.-H. (2000). Pairwise comparisons of the means of skewed data. *J. Statist. Plann. Inf.* **88**, 59–74.

Turner, T. R. (2000). Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions. *J. R. Statist. Soc. C. Appl. Statist.* **49**, 371–384.

van Es, A. J., Helmers, R., and Hušková, M. (2000). On a crossroad of resampling plans: Bootstrapping elementary symmetric polynomials. *Statistica Neerlandica* **54**, 100–110.

van Garderen, K. J., Lee, K., and Pesarian, M. H. (2000). Cross-sectional aggregation of non-linear models. *J. Econ.* **95**, 285–331.

van Toan, N. (2000). Rate of convergence in bootstrap approximations with random sample size. *Acta Mathematica Vietnamica* **25**, 161–179.

Vasdekis, V. G. S., and Trichopoulou, A. (2000). Nonparametric estimation of individual food availability along with bootstrap confidence intervals in household budget survey. *Statist. Probab. Lett.* **46**, 337–345.

Vilar-Fernández, J. M., and González-Manteiga, W. (2000). Resampling for checking linear regression models via non-parametric regression estimation. *Comput. Statist. Data Anal.* **35**, 211–231.

Vinod, H. D. (2000). Foundations of multivariate inference using modern computers. *Lin. Alg. Appl.* **321**, 365–385.

Visser, I., Raijmakers, M. E. J., and Molenaar, P. C. M. (2000). Confidence intervals for hidden Markov model parameters. *British J. Math. Statist. Psych.* **53**, 317–327.

Wang, J, Karim, R., and Medve, R. (2000). The role of bootstrap in study design. In *ASA Proceedings of the Biopharmaceutical Section*, pp. 100–104. American Statistical Association, Alexandria, VA.

Wang, Q.-H., and Jing, B.-Y. (2000). A martingale-based bootstrap inference with censored data. *Commun. Statist. Th. Meth.* **29**, 401–415.

Wang, Z. (2000). A fixed-point method for the saddlepoint approximation quantile. *Commun. Statist. Simul. Comput.* **29**, 49–60.

Wehrens, R., Putter, H., and Buydens, L. M. C. (2000) The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems* **54**, 35–52 .

Wen, S.-H., Liu, J.-P., and Ma, M.-C. (2000). On the bootstrap sample size of difference factor and similarity factor. *J. Chin. Statist. Assoc.* **38**, 193–214.

Whang, Y.-J. (2000). Consistent bootstrap tests of parametric regression functions. *J. Econ.* **98**, 27–46.

White, H. (2000). A reality check for data snooping. *Econometrica* **68**, 1097–1126.

Wilcox, R. R., Keselman, H. J., Muska, J., and Cribbie, R. (2000). Repeated measures ANOVA: Some new results on comparing trimmed means and means. *Brit. J. Math. Statist. Psych.* **53**, 69–82.

Wilkinson, R. C., Schaalje, G. B., and Collings, B. J. (2000). Confidence intervals for the kappa parameter, with application to the semiconductor industry. *Commun. Statist. Simul. Comput.* **29**, 647–665.

Winsberg, S., and deSoete, G. (2000). A bootstrap procedure for mixture models. *Data Analysis, Classification, and Related Methods* (H. A. L. Kiers, J.-P. Rasson, P. J. F. Groenen, and M. Schader, editors), pp. 59–62. Springer-Verlag, Berlin.

Wolf-Ostermann, K. (2000). Testing for differences in location: A comparison of bootstrap methods in the small sample case. In *COMPSTAT—Proceedings in Computational Statistics, 14th Symposium* (J. G. Bethlehem and P. G. M. van der Heijden, editors), pp. 517–522. Physica-Verlag Ges.m.b.H, Heidelberg.

Wood, A. T. A. (2000). Bootstrap relative errors and sub-exponential distributions. *Bernoulli* **6**, 809–834.

Woodroof, J. (2000). Bootstrapping: As easy as 1-2-3. *J. Appl. Statist.* **27**, 509–517.

Wright, J. H. (2000). Confidence intervals for univariate impulse responses with a near unit root. *J. Bus. Econ. Statist.* **18**, 368–373.

Xu, K. (2000). Inference for generalized Gini indices using the iterated-bootstrap method. *J. Bus. Econ. Statist.* **18**, 223–227.

Yu, Q., and Shie-Shien, Y. (2000). Parametric bootstrap based inference in linear mixed effects model with AR(1) correlated. ASA Proceedings of the Biopharmaceutical Section, pp. 39–44. American Statistical Association, Alexandria, VA.

Zhou, X.-H., and Gao, S. (2000). One-sided confidence intervals for means of positively skewed distributions. *Am. Statist.* **54**, 100–104.

Zhou, X.-H., and Tu, W. (2000a). Confidence intervals for the mean of diagnostic test charge data containing zeros. *Biometrics* **56**, 1118–1125.

Zhou, X.-H., and Tu, W. (2000b). Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Comput. Statist. Data Anal.* **35**, 201–210.

Zidek, J. V., and Wang, S. X. (2000). Comment on "The estimating function bootstrap." *Can. J. Statist.* **28**, 482–485.

Zucchini, W. (2000). An introduction to model selection. *J. Math. Psych.* **44**, 41–61.

2001

Aerts, M., and Claeskens, G. (2001). Bootstrap tests for misspecified models, with application to clustered binary data. *Comput. Statist. Data Anal.* **36**, 383–401.

Ahmed, S. E., Li, D., Rosalsky, A., and Voladin, A. (2001). Almost sure lim sup behavior of bootstrapped means with appleications to pairwise i.i.d. seqences and stationary ergodic sequences. *J. Statist. Plann. Inf.* **98**, 1–14.

Alba, M. V., Jimenez, M. D., Barrera, D., and Jiménez, M. D. (2001). A homogeneity test based on empirical characteristic functions. *Comput. Statist.* **16**, 255–270.

Albers, W. (2001). From A to Z: Asymptotic expansions by van Zwet. *State of the Art in Probability and Statistics*, pp. 2–20. Institute of Mathematical Statistics, Hayward.

Almudevar, A. (2001). A bootstrap assessment of variability in pedigree reconstruction based on genetic markers. *Biometrics* **57**, 757–763.

Aminzadeh, M. S. (2001). Bootstrap tolerance and confidence limits for two-variable reliability using independent and weakly dependent observations. *Appl. Math. Comput.* **122**, 81–93.

Andersson, M. K., and Karlsson, S. (2001). Bootstrapping error component models. *Comput. Statist.* **16**, 221–231.

Andrews, D. W. K., and Buchinsky, M. (2001). Evaluation of a three-step method for choosing the number of bootstrap repetitions. *J. Econ.* **103**, 345–386.

Angelova, D. S., Semerdjiev, Tz. A., Jilkov, V. P., and Semerdjiev, E. A. (2001). Application of a Monte Carlo method for tracking maneuvering target in clutter. *Math. Comput. Simul.* **55**, 15–23.

Baglivo, J. A. (2001). Teaching permutation and bootstrap methods. In *ASA Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

Bernard, A. J., and Wludyka, P. S. (2001). Robust I-sample analysis of means type randomization tests for variances. *J. Statist. Comput. Simul.* **69**, 57–88.

Bertail, P., and Politis, D. N. (2001). Extrapolation of subsampling distribution estimators: The i.i.d. and strong mixing cases. *Canad. J. Statist.* **29**, 667–680.

Bickel, P. J., and Ren, J.-J. (2001). The bootstrap in hypothesis testing. *State of the Art in Probability and Statistics*, pp. 91–112. Institute of Mathematical Statistics, Hayward.

Bisaglia, L., and Grigoletto, M. (2001). Prediction intervals for FARIMA processes by bootstrap methods. *J. Statist. Comput. Simul.* **68**, 185–201.

Bjørnstad, O. N., and Falck, W. (2001). Nonparametric spatial covariance functions: Estimation and testing. *Environ. Ecol. Statist.* **8**, 53–70.

Bloch, D. A., Lai, T. L., and Tubert-Bitter, P. (2001). One-sided tests in clinical trials with multiple endpoints. *Biometrics* **57**, 1039–1047.

Bose, A., and Chatterjee, S. (2001). Generalised bootstrap in non-regular *M*-estimation problems. *Statist. Probab. Lett.* **55**, 319–328.

Boyce, M. S., MacKenzie, D. I., Manly, B. F. J. H. M. A., and Moody, D. (2001). Negative binomial models for abundance estimation of multiple closed populations. *J. Wildlife Manag.* **65**, 498–509.

Braun, W. J., and Hall, P. (2001). Data sharpening for nonparametric inference subject to constraints. *J. Comput. Graph. Statist.* **10**, 786–806.

Brownstone, D., and Valletta, R. (2001). The bootstrap and multiple imputations: Harnessing increased computing power for improved statistical tests. *J. Econ. Perspectives* **15**, 129–141.

Bustami, R., van der Heijden, P., van Houwelingen, H., and Engbersen, G. (2001). Point and interval estimation of the population size using the truncated Poisson regression model. In *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling, Denmark 2001* (B. Klein and L. Korsholm, editors), pp. 87–94. University of Southern Denmark, Denmark.

Candelon, B., and Lütkepohl, H. (2001). On the reliability of Chow-type tests for parameter constancy in multivariate dynamic models. *Econ. Lett.* **73**, 155–160.

Caner, M., and Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica* **69**, 1555–1596.

Cantú, S. M., Villaseñor, A. J. A., and Arnold, B. C. (2001). Modeling the lifetime of longitudinal elements. *Commun. Statist. Simul. Comput.* **30**, 717–741.

Carpenter, M., and Mishra, S. N. (2001). Bootstrap bias adjusted estimators of beta distribution parameters. Calcutta Statist. *Assoc. Bull.* **51**, 119–124.

Chan, K. Y. F., and Lee, S. M. S. (2001). An exact iterated bootstrap algorithm for small-sample bias reduction. *Comput. Statist. Data Anal.* **36**, 1–13.

Chao, M. T., and Fuh, C. D. (2001). Bootstrap methods for the up and down test on pyrotechnics sensitivity analysis. *Statist. Sin.* **11**, 1–21.

Chen, C.-F., Hart, J. D., and Wang, S. (2001) Bootstrapping the order selection test. *J. Nonparam. Statist.* **13**, 851–882.

Chen, H., and Chen, J. (2001). The likelihood ratio test for homogeneity in finite mixture models. *Can. J. Statist.* **29**, 201–215.

Cheng, R. C. H., and Liu, W. B. (2001). The consistency of estimators in finite mixture models. *Scand. J. Statist.* **28**, 603–616.

Chu, K. K., Wang, N., Stanley, S., and Cohen, N. D. (2001). Statistical evaluation of the regulatory guidelines for use of furosemide in race horses. *Biometrics* **57**, 294–301.

Clark, J., Horváth, L., and Lewis, M. (2001). On the estimation of spread rate for a biological population. *Statist. Probab. Lett.* **51**, 225–234.

Clements, M. P., and Taylor, N. (2001). Bootstrapping prediction intervals for autoregressive models. *Int. J. Forecast.* **17**, 247–267.

Conversano, C., Mola, F., and Siciliano, R. (2001). Partitioning algorithms and combined model integration for data mining. *Comput. Statist.* **16**, 323–339.

Costa-Bouzas, J., Takkouche, B., Cadarso-Suárez, C., and Spiegelman, D. (2001). HEpiMA: Software for the identification of heterogeneity in meta-analysis. *Comput. Methods Prog. Biomed*. **64**, 101–107.

Cribari-Neto, F., and Zarkos, S. G. (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. *J. Statist. Comput. Simul.* **68**, 391–411.

Csörgo, S., Valkó, B., and Wu, W. B. (2001). Random multisets and bootstrap means. *Acta Sci. Math.* **67**, 843–875.

Cuevas, A., Febrero, M., and Fraiman, R. (2001). Cluster analysis: A further approach based on density estimation. *Comput. Statist. Data Anal.* **36**, 441–459.

Czado, C., and Munk, A. (2001). Bootstrap methods for the nonparametric assessment of population bioequivalence and similarity of distributions. *J. Statist. Comput. Simul.* **68**, 243–280.*

Dalrymple, M. L., Hudson, I. L., and Barnett, A. G. (2001). Survival, block bootstrap and mixture methods for detecting change points in discrete time series data with application to SIDS. In *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling, Denmark 2001* (B. Klein and L. Korsholm, editors), pp. 135–145. University of Southern Denmark, Denmark.

Delgado, M. A., and Manteiga, W. G. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.* **29**, 1469–1507.

Dette, H., and Neumeyer, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.* **29**, 1361–1400.

Diaz-Insua, M., and Rao, J. S. (2001). Mammographic computer-aided detection using bootstrap aggregation. In *ASA Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

DiCiccio, T. J., Martin, M. A., and Stern, S. E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios. *Can. J. Statist.* **29**, 67–76.

Do, K.-A., Wang, X., and Broom, B. M. (2001). Importance bootstrap resampling for proportional hazards regression. *Commun. Statist. Th. Meth.* **30**, 2173–2188.

Dupuis, D. J. (2001). Fitting log-*F* models robustly, with an application to the analysis of extreme values. *Comput. Statist. Data Anal.* **35**, 321–333.

El Bantli, F., and Hallin, M. (2001). Asymptotic behaviour of *M*-estimators in AR(p) models under nonstandard conditions. *Can. J. Statist.* **29**, 155–168.

Fercheluc, O. (2001). On the bootstrapping heteroscedastic regression models. *Probab. Math. Statist.* **21**, 265–276.

Ferreira, E., Núñez-Antón, V., and Orbe, J. (2001). A partial censored regression model in corporate finance. In *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling, Denmark 2001* (B. Klein and L. Korsholm, editors), pp. 187–190. University of Southern Denmark, Denmark.

Francis, R. I. C. C., and Manly, B. F. J. (2001). Bootstrap calibration to improve the reliability of tests to compare sample means and variances. *EnvironMetrics* **12**, 713–729.

Galindo, C. D., Liang, H., Kauermann, G., and Carroll, R. J. (2001). Bootstrap confidence intervals for local likelihood, local estimating equations and varying coefficient models. *Statist. Sin.* **11**, 121–134.

Garren, S. T., Smith, R. L., and Piergorsch, W. W. (2001). Bootstrap goodness-of-fit test for the beta-binomial model. *J. Appl. Statist.* **28**, 561–571.

Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29**, 1264–1280.

Gill, P. S., and Swartz, T. B. (2001). Statistical analyses for round robin interaction data. *Can. J. Statist.* **29**, 321–331.

Gomes, M. I., and Oliveira, O. (2001). The bootstrap methodology in statistics of extremes—Choice of the optimal sample fraction. *Extremes* **4**, 331–358.

Götze, F., and Ra kauskas, A. (2001). Adaptive choice of bootstrap sample sizes. *State of the Art in Probability and Statistics*, pp. 286–309. Institute of Mathematical Statistics, Hayward.

Greenland, S. (2001). Estimation of population attributable fractions from fitted incidence ratios and exposure survey data, with an application to electromagnetic fields and childhood leukemia. *Biometrics* **57**, 182–188.

Guillou, A., and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *J. R. Statist. Soc.* **B 63**, 293–305.

Habing, B. (2001). Nonparametric regression and the parametric bootstrap for local dependence assessment. *Appl. Psych. Meas.* **25**, 221–233.

Hall, P., and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* **29**, 624–647.

Hall, P., Huang, L. S., Gifford, J. A., and Gijbels, I. (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *J. Comput. Graph. Statist.* **10**, 592–614.

Hall, P., and Kang, K.-H. (2001). Bootstrapping nonparametric density estimators with empirically chosen bandwidths. *Ann. Statist.* **29**, 1443–1468.

Hall, P., Melville, G., and Welsh, A. H. (2001). Bias correction and bootstrap methods for a spatial sampling scheme. *Bernoulli* **7**, 829–846.

Hall, P., and Rieck, A. (2001). Improving coverage accuracy of nonparametric prediction intervals. *J. R. Statist. Soc.* **B 63**, 717–725.

Hall, P., and York, M. (2001). On the calibration of Silverman's test for multimodality. *Statist. Sin.* **11**, 515–536.

Harezlak, J., and Heckman, N. E. (2001). CriSP: A tool for bump hunting. *J. Comput. Graph. Statist.* **10**, 713–729.

Hens, N., Aerts, M., Claeskens, G., and Molenberghs, G. (2001). Multiple nonparametric bootstrap imputation. In *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling, Denmark 2001* (B. Klein and L. Korsholm, editors), pp. 219–225. University of Southern Denmark, Denmark.

Hesterberg, T. C. (2001). Bootstrap tilting diagnostics. In *ASA Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

Ho, T.-W. (2001). Finite-sample properties of the bootstrap estimator in a Markov-switching model. *J. Appl. Statist.* **28**, 835–842.

Horowitz, J. L. (2001a). The bootstrap and hypothesis tests in econometrics. *J. Econ.* **100**, 37–40.

Horowitz, J. L. (2001b). Reply to comment on "The bootstrap and hypothesis tests in econometrics". *J. Econ.* **100**, 97–98.

Hsiao, C., and Li, Q. (2001). A consistent test for conditional heteroskedasticity in time-series regression models. *Econ. Th.* **17**, 188–221.

Hu, F. (2001). Efficiency and robustness of a resampling *M*-estimator in the linear model. *J. Mult. Anal.* **78**, 252–271.

Huang, L.-S. (2001). Testing the adequacy of a linear model critical smoothing. *J. Statist. Comput. Simul.* **68**, 281–294.

Huang, Y. (2001a). Interval estimation of ED_{50} when a logistic dose–response curve is incorrectly assumed. Comput. *Statist. Data Anal.* **36**, 525–537.

Huang, Y. (2001b). Correction to: "Interval estimation of the 90% effective dose: A comparison of bootstrap resampling methods with some large-sample approaches." *J. Appl. Statist.* **28**, 516.

Huang, Y. (2001c). Various methods of interval estimation of the median effective dose. *Commun. Statist. Simul. Comput.* **30**, 99–112.

Huh, M.-H., and Jhun, M. (2001). Random permutation testing in multiple linear regression. *Commun. Statist. Th. Meth.* **30**, 2023–2032.

Hutson, A. D. (2001). Rational spline estimators of the quantile function. *Commun. Statist. Simul. Comput.* **30**, 377–390.

Hwang, Y.-T. (2001). Edgeworth expansions for the product-limit estimator under left-truncation and right-censoring with the bootstrap. *Statist. Sin.* **11**, 1069–1079.

Inoue, A., and Shintani, M. (2001). *Bootstrapping GMM estimators for time series*. Working paper, Department of Agriculture and Resource Economics, North Carolina State University.

Jackson, G., and Cheng, Y. W. (2001). Parameter estimation with egg production surveys to estimate snapper, biomass in Shark Bay, Western Australia. *J. Agr. Biol. Environ. Statist.* **6**, 243–257.

Janssen, P., Swanepoel, J., and Veraverbeke, N. (2001). Modified bootstrap consistency rates for *U*-quantiles. *Statist. Probab. Lett.* **54**, 261–268.

Jeong, J., and Chung, S. (2001). Bootstrap tests for autocorrelation. *Comput. Statist. Data Anal.* **38**, 49–69.

Kato, B. S., and Hoijtink, H. (2001). Asymptotic, Bayesian and bootstrapped *P*-values. In *New Trends in Statistical Modelling: Proceedings of the 16th International*

Workshop on Statistical Modelling, Denmark 2001 (B. Klein and L. Korsholm, editors), pp. 469–472. University of Southern Denmark, Denmark.

Kilian, L. (2001). Impulse response analysis in vector autoregressions with unknown lag order. *J. Forecast.* **20**, 161–179.

Kim, C., Hong, C., and Jeong, M. (2001). Testing the goodness of fit of a parametric model via smoothing parameter estimate. *J. Kor. Statist. Soc.* **30**, 645–660.

Kim, S., and Lee, Y.-G. (2001). Bootstrapping for ARCH models. In *ASA Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

Kim, T.-Y., Shin, K.-D., and Song, G.-M. (2001). Validity of blockwise bootstrapped empirical process with multivariate stationary sequences. *J. Kor. Statist. Soc.* **30**, 407–418.

Krzanowski, W. J. (2001). Data-based interval estimation of classification error rates. *J. Appl. Statist.* **28**, 585–595.

Lahiri, S. N. (2001). Effects of block lengths on the validity of block resampling methods. *Probab. Th. Rel. Fields* **121**, 73–97.

La Rocca, M., and Vitale, C. (2001). Parametric bootstrap inference in bilinear models. *Metron* **59**, 101–116.

Lee, K.-W., and Kim, W.-C. (2001). Bootstrap inference on the Poisson rates for grouped data. *J. Kor. Statist. Soc.* **30**, 1–20.

Lee, T.-H., and Ullah, A. (2001). Nonparametric bootstrap tests for neglected nonlinearity in time series regression models. *J. Nonparam. Statist.* **13**, 425–451.

LePage, R., and Ryznar, M. (2001). Conditioning vs. standardization for contrasts on errors attracted to stable laws. *Commun. Statist. Th. Meth.* **30**, 1829–1850.

Li, G., and Datta, S. (2001). A bootstrap approach to nonparametric regression for right censored data. *Ann. Inst. Statist. Math.* **53**, 708–729.

Liu, Q., Jin, P.-H., Gao, E.-S., and Hsieh, C.-C. (2001). Selection of prognostic factors using a bootstrap method. In *ASA Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

Lobato, I. N. (2001). *Testing that a dependent process is uncorrelated*. J. Am. Statist. Assoc. **96**, 1066–1076.

Manly, B. F. J., and Schmutz, J. A. (2001). Estimation of brood and nest survival: Comparative methods in the presence of heterogeneity. *J. Wildlife Manag.* **65**, 258–270.

Mantalos, P., and Shukur, G. (2001). Bootstrapped Johansen tests for cointegration relationships: A graphical analysis. *J. Statist. Comput. Simul.* **68**, 351–371.

Mason, D. M., and Shao, Q.-M. (2001). Bootstrapping the Student-statistic. *Ann. Probab.* **29**, 1435–1450.

Myers, L., and Hsueh, Y.-H. (2001). Using nonparametric bootstrapping to assess Kappa statistics. In *ASA Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

Nam, K. H., Kim, D. K., and Park, D. H. (2001). Large-sample interval estimators for process capability indices. *Quality Eng.* **14**, 213–221.

Naranjo, J. D., and McKean, J. W. (2001). Adjusting for regression effect in uncontrolled studies. *Biometrics* **57**, 178–181.

Nielsen, H. A., and Madsen, H. (2001). A generalization of some classical time series tools. *Comput. Statist. Data Anal.* **37**, 13–31.

Norris, J. L., III, and Pollock, K. H. (2001). Nonparametric MLE incorporation of heterogeneity and model testing into premarked cohort studies. *Environ. Eco. Statist.* **8**, 21–32.

Orbe, J. (2001). Life time analysis using a semiparametric generalized linear model. *Qüestiió* **25**, 337–358.

Öztürk, Ö. (2001). A nonparametric test of symmetry versus asymmetry for ranked-set samples. *Commun. Statist. Th. Meth.* **30**, 2117–2133.

Pan, W. (2001). Model selection in estimating equations. *Biometrics* **57**, 529–534.

Pan, W., and Le, C. T. (2001). Bootstrap model selection in generalized linear models. *J. Agr. Biol. Environ. Statist.* **6**, 49–61.

Paparoditis, E., and Politis, D. N. (2001a). Tapered block bootstrap. *Biometrika* **88**, 1105–1119.

Paparoditis, E., and Politis, D. N. (2001b). A Markovian local resampling scheme for nonparametric estimators in time series analysis. *Econ. Theory* **17**, 540–566.

Park, E., and Lee, Y. J. (2001). Estimates of standard deviation of Spearman's rank correlation coefficients with dependent observations. *Commun. Statist. Simul. Comput.* **30**, 129–142.

Pascual, L., Romo, J., and Ruiz, E. (2001a). Effects of parameter estimation on prediction densities: A bootstrap approach. *Int. J. Forecast.* **17**, 83–103.

Peddada, S. D., Prescott, K. E., and Conaway, M. (2001). Tests for order restrictions in binary data. *Biometrics* **57**, 1219–1227.

Pigeot, I. (2001). The jackknife and bootstrap in biomedical research—Common principles and possible pitfalls. *Drug Information J.* **35**, 1431–1443.*

Piraux, F., and Palm, R. (2001). Empirical study of estimators of the error levels in discriminant analysis. *Rev. Statist. Appl.* **49**, 71–85.

Polansky, A. M. (2001). Bandwidth selection for the smoothed bootstrap percentile method. *Comput. Statist. Data Anal.* **36**, 333–349.

Psaradakis, Z. (2001). Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors. *J. Time Series Anal.* **22**, 577–594.

Qin, J., and Wang, M.-C. (2001). Semiparametric analysis of truncated data. *Lifetime Data Anal.* **7**, 225–242.

Rasekh, A. R. (2001). Ridge estimation in functional measurement error models. *Annales de l'I.S.U.P.* **45**, 47–59.

Ren, J.-J. (2001). Weighted empirical likelihood ratio confidence intervals for the mean with censored data. *Ann. Inst. Statist. Math.* **53**, 498–516.

Rodríguez, G., and Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: A case study. *J. R. Statist. Soc. A* **164**, 339–355.

Shao, Y. (2001). Rate of convergence of bootstrapped empirical measures. *Statist. Probab. Lett.* **53**, 293–298.

Saavedra, P. J. (2001). An extension of Fay's method for variance estimation to the bootstrap. In *ASA Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandria, VA.

Saigo, H., Shao, T., and Sitter, R. R. (2001). A repeated half-sample bootstrap and balanced repeated replications for randomly imputed data. *Surv. Methodol.* **27**, 189–196.

Shafii, B., and Price, W. J. (2001). Estimation of cardinal temperatures in germination data analysis. *J. Agr. Biol. Environ. Statist.* **6**, 356–366.

Shoemaker, O. J., and Pathak, P. K. (2001) The sequential bootstrap: A comparison with regular bootstrap. *Commun. Statist. Th. Meth.* **30**, 1661–1674.

Simar, L., and Wilson, P. W. (2001). Testing restrictions in nonparametric efficiency models. *Commun. Statist. Simul. Comput.* **30**, 159–184.

Sjöstedt-de Luna, S. (2001). Resampling non-homogeneous spatial data with smoothly varying mean values. *Statist. Probab. Lett.* **53**, 373–379.

Smith, W. D., and Taylor, R. L. (2001). Consistency of dependent bootstrap estimators. *Am. J. Math. Manag. Sci.* **21**, 359–382.

Solow, A. R., and Costello, C. J. (2001). A test for declining diversity. *Ecology* **82**, 2370–2372.

Souza, M., Andréia A., and Louzada-Neto, F. (2001). Inference for parameters of the bi-log-logistic model. *Rev. Mat. Estatist.* **19**, 309–324.

Sullivan, R., Timmermann, A., and White, H. (2001). Dangers of data mining: The case of calendar effects in stock returns. *J. Econ.* **105**, 249–286.

Sun, Y., Sun, S., and Diao, Y. (2001). Smooth quantile processes from right censored data and construction of simultaneous confidence bands. *Commun. Statist. Th. Meth.* **30**, 707–727.

Traat, I., Meister, K., and Söstra, K. (2001). Statistical inference in sampling theory. *Theory Stoch. Proc.* **7**, 301–316.

van der Laan, M. J., and Bryan, J. (2001). Gene expression analysis with the parametric bootstrap. *Biostatistics* **2**, 445–461.

Velilla, S. (2001). On the bootstrap in misspecified regression models. *Comput. Statist. Data Anal.* **36**, 227–242.

Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Appl. Psych. Meas.* **25**, 283–294.

Wellner, J. A. (2001). Some converse limit theorems for exchangeable bootstraps. *State of the Art in Probability and Statistics*, pp. 593–606. Institute of Mathematical Statistics, Hayward.

Wilcox, R. R. (2001). Pairwise comparisons of trimmed means for two or more groups. *Psychometrika* **66**, 343–356.

Wood, G. C., Frey, M. R., and Frey, C. M. (2001). A bootstrap test of the goodness-of-fit of the multivariate tolerance model. *Commun. Statist. Simul. Comput.* **30**, 477–488.

Wood, S. N. (2001). Minimizing model fitting objectives that contain spurious local minima by bootstrap restarting. *Biometrics* **57**, 240–244.

Yanez, N. D., III, Warnes, G. R., and Kronmal, R. A. (2001). A univariate measurement error model for longitudinal change. *Commun. Statist. Th. Meth.* **30**, 279–287.

Yu, K., and Feingold, E. (2001). Estimating the frequency distribution of crossovers during meiosis from recombination data. *Biometrics* **57**, 427–434.

Yue, J. C., Clayton, M. K., and Lin, F.-C. (2001). A nonparametr estimator of species overlap. *Biometrics* **57**, 743–749.

Zhang, D. (2001). Bayesian bootstraps for *U*-processes, hypothesis tests and convergence of Dirichlet *U*-processes. *Statist. Sin.* **11**, 463–478.

Ziegler, K. (2001). On bootstrapping the mode in the nonparametric regression model with random design. *Metrika* **53**, 141–170.

Zhu, J., and Lahiri, S. N. (2001). Weak convergence of blockwise bootstrap empirical processes for stationary random fields with statistical applications. *Preprint*, Department of Statistics, Iowa State University.

2002

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *J. Am. Statist. Assoc.* **97**, 284–292.

Aerts, M., Claeskens, G., Hens, N., and Molenberghs, G. (2002). Local multiple imputation. *Biometrika* **89**, 375–388.

Alonso, A. M., Peña, D., and Romo, J. (2002). Forecasting time series with sieve bootstrap. *J. Statist. Plann. Inf.* **100**, 1–11.

Andrews, D. W. K. (2002). Higher-order improvements of a computationally attractive *k*-step bootstrap for extremum estimators. *Econometrica* **70**, 119–162.

Andrews, D. W. K., and Buchinsky, M. (2002). On the number of bootstrap repetitions for BC_a confidence intervals. *Econ. Th.* **18**, 962–984.

Astatkie, T. K., Yiridoe, E., and Clark, J. S. (2002). Testing for trend in the variability of monthly and seasonal temperature using the range and the standard deviation as measures of variability. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 79–83, American Statistical Association, Alexandria, VA.

Babu, G. J., and Padmanabhan, A. R. (2002). Resampling methods for the nonparametric Behrens–Fisher problem. *Sankhy A* **64**, 678–692.

Benton, D., and Krishnamoorthy, K. (2002). Performance of the parametric bootstrap method in small sample interval estimates. *Adv. Appl. Statist.* **2**, 269–285.

Bhattacharya, R., and Patrangenaru, V. (2002). Nonparametic estimation of location and dispersion on Riemannian manifolds. *J. Statist. Plann. Inf.* **108**, 23–35.

Bickel, P. J., and Sakov, A. (2002). Extrapolation and the bootstrap. *Sankhyā A* **64**, 640–652.

Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *J. Econ.* **108**, 317–342.

Bilder, C. R., and Loughin, T. M. (2002). Testing for conditional multiple marginal independence. *Biometrics* **58**, 200–208.

Bloch, D. A., Olshen, R. A., and Walker, M. G. (2002). Risk estimation for classification trees. *J. Comput. Graph. Statist.* **11**, 263–288.

Buonaccorsi, J. P., and Elkinton, J. S. (2002). Regression analysis in a spatial–temporal context: Least squares, generalized least squares, and the use of the bootstrap. *J. Agr. Biol. Environ. Statist.* **7**, 4–20.

Brailsford, T. J., Penm, J. H. W., and Terrell, R. D. (2002). Selecting the forgetting factor in subset autoregressive modeling. *J. Time Series Anal.* **23**, 629–649.

Brown, B. M. (2002). On the importance of being smooth. *Aust. New Zealand J. Statist.* **44**, 143–154.

Brown, B. W., and Newey, W. K. (2002). Generalized method of moments, efficient bootstrapping, and improved inference. *J. Bus. Econ. Statist.* **20**, 507–517.

Bühlmann, P. (2002a). Sieve bootstrap with variable-length Markov chains for stationary categorical time series. *J. Am. Statist. Assoc.* **97**, 443–456.*

Bühlmann, P. (2002b). Reply to comments on "Sieve bootstrap with variable-length Markov chains for stationary categorical time series." *J. Am. Statist. Assoc.* **97**, 466–471.

Bühlmann, P. (2002c). Bootstraps for time series. *Statisti. Sci.* **17**, 52–72.

Bühlmann, P., and Yu, B. (2002). Analyzing bagging. *Ann. Statist.* **30**, 927–961.

Butar, F. B., and Lahiri, P. (2002). Empirical Bayes estimation of several population means and variances under random sampling variances model. *J. Statist. Plann. Inf.* **102**, 59–69.

Butler, R. W., and Bronson, D. A. (2002). Bootstrapping survival times in stochastic systems by using saddlepoint approximations. *J. R. Statist. Soc. B.* **64**, 31–49.

Butler, R. W., and Paolella, M. S. (2002). Saddlepoint approximation and bootstrap inference for the Satterthwaite class of ratios. *J. Am. Statist. Assoc.* **97**, 836–846.

Carpenter, M. (2002). Estimation of location extremes within general families of scale mixtures. *J. Statist. Plann. Inf.* **100**, 197–208.

Chatterjee, S., and Bose, A. (2002). Dimension asymptotics for generalised bootstrap in linear regression. *Ann. Inst. Statist. Math.* **54**, 367–381.

Chaubey, Y. P. (2002). Estimation in inverse Gaussian regression: Comparison of asymptotic and bootstrap distributions. *J. Statist. Plann. Inf.* **100**, 135–143.

Chen, J. J., and Wang, S.-J. (2002). Testing for treatment effects on subsets of endpoints. *Biomed. J.* **44**, 541–557.

Chen, J.-P. (2002). Evaluating the bootstrap p control chart. *Adv. Appl. Statist.* **2**, 19–30.

Chen, S. X., Yip, P. S. F., and Zhou, Y. (2002). Sequential estimation in line transect surveys. *Biometrics* **58**, 263–269.

Chernick, M.R. and Friis, R.H. (2002). *Introductory Biostatistics for the Health Sciences Modern Applications Including Bootstrap*. Wiley, New York.*

Climov, D., Delecroix, M., and Simar, L. (2002). Semiparametric estimation in single index Poisson regression: A practical approach. *J. Appl. Statist.* **29**, 1047–1070.

Corcoran, C. D., and Mehta, C. R. (2002). Exact level and power of permutation, bootstrap, and asymptotic tests of trend. *J. Mod. Appl. Statist. Meth.* **1**, 42–51.

Conversano, C., Siciliano, R., and Mola, F. (2002). Generalized additive multi-mixture model for data mining. *Comput. Statist. Data Anal.* **38**, 487–500.

Cribari-Neto, F., Frery, A. C., and Silva, M. F. (2002). Improved estimation of clutter properties in speckled imagery. *Comput. Statist. Data Anal.* **40**, 801–824.

D'Alessandro, L., and Fattorini, L. (2002). Resampling estimators of species richness from presence–absence data: Why they don't work? *Metron* **60**, 5–19.

Damien, P., and Walker, S. (2002). A Bayesian non-parametric comparison of two treatments. *Scand. J. Statist.* **29**, 51–56.

Davidson, J. (2002). A model of fractional cointegration and tests for cointegration, using the bootstrap. *J. Econ.* **110**, 187–212.

Davidson, R., and MacKinnon, J. G. (2002a). Bootstrap *J* tests of nonnested linear regression models. *J. Econ.* **109**, 167–193.

Davidson, R., and MacKinnon, J. G. (2002b). Fast double bootstrap tests of nonnested linear regression models. *Econ. Rev.* **21**, 419–429.

Delgado, M. A., and Fiteni, I. (2002). External bootstrap tests for parameter stability. *J. Econ.* **109**, 275–303.

Demirel, O. F., and Willemain, T. R. (2002). A Turing test of bootstrap scenarios. *J. Comput. Graph. Statist.* **11**, 896–909.

DiCiccio, T. J., and Monti, A. C. (2002). Accurate confidence limits for scalar functions of vector—estimands. *Biometrika* **89**, 437–450.

Dmitrienko, A., and Govindarajulu, Z. (2002). Sequential determination of the number of bootstrap samples. *J. Statist. Plann. Inf.* **100**, 349–363.

Dufour, J.-M., and Khalaf, L. (2002a). Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions. *J. Econ.* **106**, 143–170.

Dufour, J.-M., and Khalaf, L. (2002b). Simulation based finite and large sample tests in multivariate regressions. *J. Econ.* **111**, 303–322.

Eberhardt, L. L. (2002). A paradigm for population analysis of long-lived vertebrates. *Ecology* **83**, 2841–2854.

Efron, B. (2002). The bootstrap and modern statistics. In *Statistics in the 21st Century* (A. E. Raftery, M. A. Tanner, and M. Wells, editors), pp. 326–332. Chapman & Hall, London.

Fan, Y., and Li, Q. (2002). A consistent model specification test based on the kernel sum of squares of residuals. *Econ. Rev.* **21**, 337–352.

Fay, M. P., and Follmann, D. A. (2002). Designing Monte Carlo implementations of permutation or bootstrap hypothesis tests. *Am. Statist.* **56**, 63–70.

Fernholz, L. T. (2002a). On smoothing the corrected content of tolerance intervals. *Estadística* **54**, 91–126.

Fernholz, L. T. (2002b). Robustness issue regarding content-corrected tolerance limits. *Metrika* **55**, 53–66.

Ferreira, A. (2002). Optimal asymptotic estimation of small exceedance probabilities. *J. Statist. Plann. Inf.* **104**, 83–102.

Flachaire, E. (2002). Bootstrapping heteroskedasticity consistent covariance matrix estimator. *Comput. Statist.* **17**, 501–506.

Formann, A. K., and Ponocny, I. (2002). Latent change classes in dichotomous data. *Psychometrika* **67**, 437–457.

Franco, G. C., and Souza, R. C. (2002). A comparison of methods for bootstrapping in the local level model. *J. Forecast.* **21**, 27–38.

Franke, J., Kreiss, J.-P., and Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli.* **8**, 1–37.

Franke, J., Kreiss, J.-P., Mammen, E., and Neumann, M. H. (2002). Properties of the nonparametric autoregressive bootstrap. *J. Time Series Anal.* **23**, 555–585.

Friedl, H., and Stampfer, E. (2002). Estimating general variable acceptance sampling plans by bootstrap methods. *J. Appl. Statist.* **29**, 1205–1217.

Gavaris, S., and Ianelli, J. N. (2002). Statistical issues in fisheries' stock assessments. *Scand. J. Statist.* **29**, 245–271.

Ginevan, M. E., and Splitstone, D. E. (2002). Bootstrap upper bounds for the arithmetic mean of right-skewed data, and the use of censored data. *EnvironMetrics* **13**, 453–464.

Gombay, E., and Horváth, L. (2002). Rates of convergence for *U*-statistic processes and their bootstrapped versions. *J. Statist. Plann. Inf.* **102**, 247–272.

Gonçalves, S., and White, H. (2002). The bootstrap of the mean for dependent heterogeneous arrays. *Econ. Th.* **18**, 1367–1384.

Gospodinov, N. (2002a). Bootstrap-based inference in models with a nearly noninvertible moving average component. *J. Bus. Econ. Statist.* **20**, 254–268.

Gospondinov, N. (2002b). Median unbiased forecasts for highly persistent autoregressive processes. *J. Econ.* **111**, 85–101.

Hall, P. (2002). Comment on "Sieve bootstrap with variable-length Markov chains for stationary categorical time series." *J. Am. Statist. Assoc.* **97**, 456–457.

Hall, P., Peng, L., and Tajvidi, N. (2002). Effect of extrapolation on coverage accuracy of prediction intervals computed from Pareto-type data. *Ann. Statist.* **30**, 875–895.

Hall, P., and Tajvidi, N. (2002). Permutation tests for equality of diributions in high-dimensional settings. *Biometrika* **89**, 359–374.

Hall, P., and Park, B. U. (2002). New methods for bias correction at endpoints and boundaries. *Ann. Statist.* **30**, 1460–1479.

Hall, P., Peng, L., and Tajvidi, N. (2002). Effect of extrapolation on coverage accuracy of prediction intervals computed from Pareto-type data. *Ann. Statist.* **30**, 875–895.

Hall, P., Peng, L., and Yao, Q. (2002). Moving-maximum models for extrema of time series. *J. Statist. Plann. Inf.* **103**, 51–63.

Hansen, B. E., and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *J. Econ.* **110**, 293–318.

Hansen, B. E., and West, K. D. (2002). Generalized method of moments and macroeconomics. *J. Bus. Econ. Statist.* **20**, 460–469.

He, X., and Hu, F. (2002). Markov chain marginal bootstrap. *J. Am. Statist. Assoc.* **97**, 783–795.

Henze, N., and Klar, B. (2002). Goodness-of-fit tests for the inverse Gaussian distribution based on the empirical Laplace transform. *Ann. Inst. Statist. Math.* **54**, 425–444.

Hesterberg, T. (2002). Bootstrap tilting diagnostics. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 1435–1438, American Statistical Association, Alexandria, VA.

Horowitz, J. L. (2002). Bootstrap critical values for tests based on the smoothed maximum score estimator. *J. Econ.* **111**, 141–167.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–128.

Huang, Y. (2002). Robustness of interval estimation of the 90% effective dose: Bootstrap resampling and some large-sample parametric methods. *J. Appl. Statist.* **29**, 1071–1081.

Huggins, R. M. (2002). A parametric empirical Bayes approach to the analysis of capture–recapture experiments. *Aust. New Zealand J. Statist.* **44**, 55–62.

Hutson, A. D. (2002). Analytical bootstrap methods for censored data. *J. Appl. Math. Dec. Sci.* **6**, 129–141.

Hutson, A. D. (2002). "Exact" bootstrap confidence bands for the quantile function via Steck's determinant. *J. Comput. Graph. Statist.* **11**, 471–482.

Ichikawa, M., and Konishi, S. (2002). Asymptotic expansions and bootstrap approximations in factor analysis. *J. Mult. Anal.* **81**, 47–66.

Inoue, A., and Kilian, L. (2002a). Bootstrapping autoregressive processes with possible unit roots. *Econometrica* **70**, 377–391.

Inoue, A., and Kilian, L. (2002b). Bootstrapping smooth functions of slope parameters and innovation variances in infinite-order VAR models. *Int. Econ. Rev.* **43**, 309–332.

Jing, B.-Y., Kolassa, J. E., and Robinson, J. (2002). Partial saddlepoint approximations for transformed means. *Scand. J. Statist.* **29**, 721–731.

Karlis, D., and Kostaki, A. (2002). Bootstrap techniques for mortality models. *Biomed. J.* **44**, 850–866.

Keselman, H. J., Wilcox, R. R., Othman, A. R., and Frandette, K. (2002). Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and nonnormality. *J. Mod. Appl. Statist. Meth.* **1**, 288–309.

Kieser, M., Schneider, B., and Friede, T. (2002). A bootstrap procedure for adaptive selection of the test statistic in flexible two-stage designs. *Biomed. J.* **44**, 641–652.

Kim, J. H. (2002). Bootstrap prediction intervals for autoregressive models of unknown or infinite lag order. *J. Forecast.* **21**, 265–280.

Kim, J. K. (2002). A note on approximate Bayesian bootstrap imputation. *Biometrika* **89**, 470–477.

Kim, Y., and Lee, S. (2002). On the Kolmogorov-Smirnov type test for testing nonlinearity in time series. *Commun. Statist. Th. Meth.* **31**, 299–309.

Kosorok, M. R., Fine, J. P., Jiang, H., and Chappell, R. (2002). Asymptotic theory for the gamma frailty model with dependent censoring. *Ann. Inst. Statist. Math.* **54**, 476–499.

Lahiri, S. N. (2002a). On the jackknife-after-bootstrap method for dependent data and its consistency properties. *Econ. Theory.* **18**, 79–98.

Lahiri, S. N. (2002b). Comment on "Sieve bootstrap with variable-length Markov chains for stationary categorical time series." *J. Am. Statist. Assoc.* **97**, 460–461.

Lahiri, S. N., Lee, Y.-D., and Cressie, N. (2002). Efficiency of least squares estimators of spatial variogram parameters. *J. Statist. Plann. Inf.* **3**, 65–85.

Lam, J.-P., and Veall, M. R. (2002). Bootstrap prediction intervals for single period regression forecasts. *Int. J. Forecast.* **18**, 125–130.

Lee, Y.-D., and Lahiri, S. N. (2002). Least squares variagram fitting by spatial subsampling. *J. R. Statist. Soc. B.* **64**, 837–854.

Li, D., and Rosalsky, A. (2002). Erdös–Rényi–Shepp laws for arrays with application to bootstrapped sample means. *Pak. J. Statist.* **18**, 255–263.

Louzada-Neto, F., and Castro Perdoná, G. (2002). Accelerated lifetime tests with a log-non-linear stress–response relationship. *Commun. Statist. Th. Meth.* **31**, 129–146.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., and Royle, J. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248–2255.

Marazzi, A. (2002). Bootstrap tests for robust means of asymmetric distributions with unequal shapes. *Comput. Statist. Data Anal.* **39**, 503–528.

Marazzi, A., and Barbati, G. (2002). Robust parametric means of asymmetric distributions: Estimation and testing. *Estadística* **54**, 47–72.

Menshikov, M. V., Rybnikov, K. A., and Volkov, S. E. (2002). The loss of tension in an infinite membrane with holes distributed according to a Poisson law. *Adv. Appl. Probab*, **34**, 292–312.

Modarres, R. (2002). Efficient nonparametric estimation of a distribution function. *Comput. Statist. Data Anal.* **39**, 75–95.

Mokhlis, N. A., and Ibrahim, S. (2002). Efficient bootstrap resampling for dependent data. *Commun. Statist. Simul. Comput.* **31**, 345–355.

Müller, I., and El-Shaarawi, A. H. (2002). Confidence intervals for the calibration estimator with environmental applications. *EnvironMetrics* **13**, 29–42.

Nze, P. A., Bühlmann, P., and Doukhan, P. (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Ann. Statist.* **30**, 397–430.

Ohman-Strickland, P., and Casella, G. (2002). Approximate and estimated saddlepoint approximations. *Canad. J. Statist.* **30**, 97–108.

Ombao, H., Raz, J., von Sachs, R., and Guo, W. (2002). The SLEX model of a non-stationary random process. *Ann. Inst. Statist. Math.* **54**, 171–200.

Pallini, A. (2002a). Simultaneous confidence bands for pair correlation functions in Markov point processes. *Quaderni di Statistica* **4**, 1–16.

Pallini, A. (2002b). Non-parametric confidence intervals for correlations in nearest-neighbour Markov point processes. *EnvironMetrics* **13**, 187–207.

Pan, W., and Chappell, R. (2002). Esting in the Cox proportional hazards model with left-truncated and interval-censored data. *Biometrics* **58**, 64–70.

Paparoditis, E., and Politis, D. N. (2002a). The tapered block bootstrap for general statistics from stationary sequences. *Econ. J. Online* **5**, 131–148.

Paparoditis, E., and Politis, D. N. (2002b). Local block bootstrap. *Comptes Rendus: Mathématique* **335**, 959–962.

Paparoditis, E., and Politis, D. N. (2002c). The local bootstrap for Markov processes. *J. Statist. Plann. Inf.* **108**, 301–328.

Park, E. S., Spiegelman, C. H., and Henry, R. C. (2002). Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models. *EnvironMetrics* **13**, 775–798.

Park, J. Y. (2002). An invariance principle for sieve bootstrap in time series. *Econ. Theory* **18**, 469–490.

Patel, H. I. (2002). Robust analysis of a mixed-effect model for a multicenter clinical trial. *J. Biopharm. Statist.* **12**, 21–37.

Patrangenaru, V., and Mardia, K. V. (2002). A bootstrap approach to Pluto's origin. *J. Appl. Statist.* **29**, 935–943.

Pewsey, A. (2002). Testing circular symmetry. *Can. J. Statist.* **30**, 591–600.

Politis, D. N. (2002). Comment on "Sieve bootstrap with variable-length Markov chains for stationary categorical time series." *J. Am. Statist. Assoc.* **97**, 463–465.

Procidano, I., and Luchini, S. R. (2002). Testing unit roots by bootstrap. *Metron* **60**, 175–189.

Roy, D., and Safiquzzaman, M. D. (2002). Jackknifing a general class of estimators—A new approach with reference to varying probability sampling. *Pak. J. Statist.* **18**, 395–414.

Rueda, C., Menéndez, J. A., and Salvador, B. (2002). Bootstrap adjusted estimators in a restricted setting. *J. Statist. Plann. Inf.* **107**, 123–131.

Runkle, D. E. (2002). Vector autoregressions and reality. *J. Bus. Econ. Statist.* **20**, 128–133.

Salicrú, M., Vives, S., Sanchez, J., and Oliva, F. (2002). Asymptotic and bootstrap methodologies on the estimation of the sonorous sensation. *Adv. Appl. Statist.* **2**, 311–322.

Samawi, H. M., Mustafa, A. S. B., and Ahmed, M. S. (2002). Importance resampling using chi-square tilting. *Metron*, **60**, 183–200.

Sandford, B. P., and Smith, S. G. (2002). Estimation of smolt-to-adult return percentages for Snake River basin anadromous salmonids, 1990–1997. *J. Agr. Biol. Environ. Statist.* **7**, 243–263.

Schweder, T., and Hjort, N. L. (2002). Confidence and likelihood. *Scand J. Statist.* **29**, 309–332.

Sherman, M. (2002). Comment on "Sieve bootstrap with variable-length Markov chains for stationary categorical time series." *J. Am. Statist. Assoc.* **97**, 457–460.

Shieh, Y.-Y., and Fouladi, R. (2002). The application of bootstrap methodology to multilevel mixed effects linear models under conditions of error term nonnormality. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 3191–3196. American Statistical Association, Alexandria, VA.

Shimodaira, H. (2002). Assessing the uncertainty of the cluster analysis using the bootstrap resampling. *Proc. Inst. Statist. Math.* **50**, 33–44.

Smith, B. M., and Gemperline, P. J. (2002). Bootstrap methods for assessing the performance of near-infrared pattern classification techniques. *J. Chemometrics* **16**, 241–246.

Smith, R. W. (2002). The use of random-model tolerance intervals in environmental monitoring and regulation. *J. Agr. Biol. Environ. Statist.* **7**, 74–94.

Spiegelman, C. H., and Henry, R. C. (2002). Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models. *EnvironMetrics* **13**, 775–798.

Stampfer, E., and Stadlober, E. (2002). Methods for estimating principal points. *Commun. Statist. Simul. Comput.* **31**, 261–277.

Swanepoel, J. W. H., and Van Graan, F. C. (2002). Goodness-of-fit tests based on estimated expectations of probability integral transformed order statistics. *Ann. Inst. Statist. Math.* **54**, 531–542.

Tambakis, D. N., and Royen, A.-S. Van (2002). Conditional predictability of daily exchange rates. *J. Forecast.* **21**, 301–315.

Tamhane, A. C., and Logan, B. R. (2002). Multiple test procedures for identifying the minimum effective and maximum safe doses of a drug. *J. Am. Statist. Assoc.* **97**, 293–301.

Tu, W., and Zhou, X.-H. (2002). A bootstrap confidence interval procedure for the treatment effect using propensity score subclassification. Health Serv. *Outcomes Res. Meth.* **3**, 135–147.

van Giersbergen, N. P. A., and Kiviet, J. F. (2002). How to implement the bootstrap in static or stable dynamic regression models: Test statistic versus confidence region approach. *J. Econ.* **108**, 133–156.

Van Keilegom, I., and Hettmansperger, T. P. (2002). Inference on multivariate *M* estimators based on bivariate censored data. *J. Am. Statist. Assoc.* **97**, 328–336.

Ventura, V., Carta, R., Kass, R. E., Gettner, S. N., and Olson, C. R. (2002). Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics* **3**, 1–20.

Vilar-Fernández, J. M., and Vilar-Fernández, J. A. (2002). Bootstrap of minimum distance estimators in regression with correlated disturbances. *J. Statist. Plann. Inf.* **108**, 283–299.

Wall, K. D., and Stoffer, D. S. (2002). A state space approach to bootstrapping conditional forecasts in ARMA models. *J. Time Series Anal.* **23**, 733–751.

Wang, H. H., and Zhang, H. (2002). Model-based clustering for cross-sectional time series data. *J. Agr. Biol. Environ. Statist.* **7**, 107–127.

Wang, Q., and Rao, J. N. K. (2002). Empirical likelihood-based inference in linear errors-in-covariables models with validation data. *Biometrika* **89**, 345–358.

Wilcox, R. R. (2002a). Comparing the variances of two independent groups. *Brit. J. Math. Statist. Psych.* **55**, 169–175.

Wilcox, R. R. (2002b). Multiple comparisons among dependent groups based on a modified one-step M-estimator. *Biomed. J.* **44**, 466–477.

Wilcox, R. R., and Keselman, H. J. (2002a). Within groups multiple comparisons based on robust measures of location. *J. Mod. Appl. Statist. Meth.* **1**, 281–287.

Wilcox, R. R., and Keselman, H. J. (2002b). Power analyses when comparing trimmed means. *J. Mod. Appl. Statist. Meth.* **1**, 24–31.

Winsberg, S., and Soete, G. De (2002). A bootstrap procedure for mixture models: Applied to multidimensional scaling latent class models. *Appl. Stoch. Mod. Bus. Ind.* **18**, 391–406.

Wong, W., and Ho, C.-C. (2002). Evaluating the effect of sample size changes on scoring system performance using bootstraps and random samples. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 3777–3782. American Statistical Association, Alexandria, VA.

Yasui, Y., Yanai, H., Sawanpanyalert, P., and Tanaka, H. (2002). A statistical method for the estimation of window-period risk of transfusion-transmitted HIV in donor screening under non-steady state. *Biostatistics* **3**, 133–143.

Yu, K., and Feingold, E. (2002). Methods for analyzing the spatial distribution of chiasmata during meiosis based on recombination data. *Biometrics* **58**, 369–377.

Yuen, K. C., Zhu, K., and Zhang, D. (2002). Comparing *k* cumulative incidence functions through resampling methods. *Lifetime Data Anal.* **8**, 401–412.

Zhang, B. (2002). Assessing goodness-of-fit of generalized logit models based on case-control data. *J. Mult. Anal.* **82**, 17–38.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39–48.

Zhu, L. X., Yuen, K. C., and Tang, N. Y. (2002). Resampling methods for testing a semiparametric random censorship model. *Scand. J. Statist.* **29**, 111–123.

2003

Almasri, A., and Shukur, G. (2003). An illustration of the causality relation between government spending and revenue using wavelet analysis on Finnish data. *J. Appl. Statist.* **30**, 571–584.

Alonso, A. M., Peña, D., and Romo, J. (2003). Resampling time series using missing values techniques. *Ann. Inst. Statist. Math.* **55**, 765–796.

Arcones, M. A. (2003). On the asymptotic accuracy of the bootstrap under arbitrary resampling size. *Ann. Inst. Statist. Math.* **55**, 563–583.

Babu, G. J., Singh, K., and Yang, Y. (2003a). Confidence limits to the distance of the true distribution from a misspecified family by bootstrap. *J. Statist. Plann. Inf.* **115**, 471–478.

Babu, G. J., Singh, K., and Yang, Y. (2003b). Edgeworth expansions for compound Poisson processes and the bootstrap. *Ann. Inst. Statist. Math.* **55**, 83–94.

Baklizi, A. (2003). Confidence intervals for P(X < Y). *J. Mod. Appl. Statist. Meth.* **2**, 341–349.

Barrett, G. F., and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica* **71**, 71–104.

Beyene, J., Hallett, D. C., and Shoukri, M. (2003). On the use of the bootstrap for statistical inference in critical care medicine. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 534–538, American Statistical Association, Alexandria, VA.

Boulier, B. L., and Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *Int. J. Forecast.* **19**, 257–270.

Braun, W. J., and Kulperger, R. J. (2003). Re-colouring the intensity-based bootstrap for point processes. *Commun. Statist. Simul. Comput.* **32**, 475–488.

Braun, W. J., Rousson, V., Simpson, W. A., and Prokop, J. (2003). Parametric modeling of reaction time experiment data. *Biometrics* **59**, 661–669.

Bretz, F., and Hothorn, L. A. (2003). Comparison of exact and resampling based multiple testing procedures. *Commun. Statist. Simul. Comput.* **32**, 461–473.

Bun, M. J. G. (2003). Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix. *Econ. Rev.* **22**, 29–58.

Butar, F. B., and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *J. Statist. Plann. Inf.* **112**, 63–76.

Cai, J., and Kim, J. (2003). Nonparametric quantile estimation with correlated failure time data. *Lifetime Data Anal.* **9**, 357–371.

Camba-Mendez, G., Kapetanios, G., Smith, R. J., and Weale, M. R. (2003). Tests of rank in reduced rank regression models. *J. Bus. Econ. Statist.* **21**, 145–155.

Canepa, A. (2003). Bartlett correction for the LR test in cointegrating models: A bootstrap approach. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 795–802. American Statistical Association, Alexandria, VA.

Carpenter, J. R., Goldstein, H., and Rasbash, J. (2003). A novel bootstrap procedure for assessing the relationship between class size and achievement. *J. R. Statist. Soc. C, Appl. Statis.* **52**, 431–443.

Casella, G. (2003). Introduction to the silver anniversary of the bootstrap. *Statist. Sci.* **18**, 133–134.

Chatterjee, S., and Das, S. (2003). Parameter estimation in conditional heteroscedastic models. *Commun. Statist. Th. Meth.* **32**, 1135–1153.

Chen, S. X., and Hall, P. (2003). Effects of bagging and bias correction on estimators defined by estimating equations. *Statist. Sin.* **13**, 97–109.

Chen, S. X., Leung, D. H. Y., and Qin, J. (2003). Information recovery in a study with surrogate endpoints. *J. Am. Statist. Assoc.* **98**, 1052–1062.

Chen, X., Linton, O., and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71**, 1591–1608.

Claeskens, G., Aerts, M., and Molenberghs, G. (2003). A quadratic bootstrap method and improved estimation in logistic regression. *Statist. Probab. Lett.* **61**, 383–394.

Claeskens, G., Jing, B.-Y., Peng, L., and Zhou, W. (2003). Empirical likelihood confidence regions for comparison distributions and ROC curves. *Can. J. Statist.* **31**, 173–190.

Claeskens, G., and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852–1884.

Corrente, J. E., Chalita, L. V. A. S., and Moreira, J. A. (2003). Choosing between Cox proportional hazards and logistic models for interval-censored data via bootstrap. *J. Appl. Statist.* **30**, 37–47.

Csörgo, S., and Rosalsky, A. (2003). A survey of limit laws for bootstrapped sums. *Int. J. Math. & Math. Sci.* **2003**, 2835–2861.

Dilleen, M., Heimann, G., and Hirsch, I. (2003). Non-parametric estimators of a monotonic dose-response curve and bootstrap confidence intervals. *Statist. Med.* **22**, 869–882.

Domínguez, M. A., and Lobato, I. N. (2003). Testing the martingale difference hypothesis. *Econ. Rev.* **22**, 351–377.

Duckworth, W. M., and Stephenson, W. R. (2003). Resampling methods: Not just for statisticians anymore. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 1280–1285. American Statistical Association, Alexandria, VA.

Durban, M., Hackett, C. A., McNicol, J. W., Newton, A. C., Thomas, W. T. B., and Currie, I. D. (2003). The practical use of semiparametric models in field trials. *J. Agr. Biol. Environ. Statist.* **8**, 48–66.

Efron, B. (2003). Second thoughts on the bootstrap. *Statist. Sci.* **18**, 135–140.

Ferro, C. A. T., and Segers, J. (2003). Inference for clusters of extreme values. *J. R. Statist. Soc. B.* **65**, 545–556.

Formann, A. K. (2003a). Latent class model diagnostics—A review and some proposals. *Comput. Statist. Data Anal.* **41**, 549–559.

Formann, A. K. (2003b). Latent class model diagnosis from a frequentist point of view. *Biometrics* **59**, 189–196.

Gao, Y. (2003). One simple test of symmetry. *J. Probab. Statist. Sci.* **1**, 129–134.

Gelman, A. (2003). A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *Int. Stat. Rev.* **71**, 369–382.

Ginevan, M. E. (2003). Bootstrap-Monte Carlo hybrid upper confidence bounds for right skewed data. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 1609–1612, American Statistical Association Alexandria, VA.

Gonçalves, S., and de Jong, R. (2003). Consistency of the stationary bootstrap under weak moment conditions. *Econ. Lett.* **81**, 273–278.

Granger, C. W. J., and Jeon, Y. (2003). Comparing forecasts of inflation using time distance. *Int. J. Forecast.* **19**, 339–349.

Guillou, A., and Merlevède, F. (2003). Second-order properties of the blocks of blocks bootstrap for density estimators for continuous time processes. *Math. Meth. Statist.* **12**, 1–30.

Gulati, S., and Neus, J. (2003) Goodness of fit statistics for the exponential distribution when the data are grouped. *Commun. Statist. Th. Meth.* **32**, 681–700.

Hall, P., and Yao, Q. (2003a). Data tilting for time series. *J. R. Statist. Soc. B* **65**, 425–442.

Hall, P., and Yao, Q. (2003b). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**, 285–317.

Hall, P., and Zhou, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* **31**, 201–224.

Halloran, M. E., Préziosi, M. P., and Chu, H. (2003). Estimating vaccine efficacy from secondary attack rates. *J. Am. Statist. Assoc.* **98**, 38–46.

Hardle, W., Horowitz, J., and Kreiss, J.-P. (2003). Bootstrap methods for time series. *Int. Statist. Rev.* **71**, 435–459.

Harvill, J. L., and Ray, B. K. (2003). Functional coefficient autoregressive modeling for multivariate temporal data. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 1772–1782, American Statistical Association, Alexandria, VA.

Heckelei, T., and Mittelhammer, R. C. (2003). Bayesian bootstrap multivariate regression. *J. Econ.* **112**, 241–264.

Hesterberg, T., Moore, D.S., Monaghan, S., Clipson, A., and Epstein, R. (2003). *Bootstrap Methods and Permutation Tests: Companion Chapter 18 to the Practice of Business Statistics*, W.H. Freeman and Company, New York.*

Holroyd, A. E. (2003). Sharp metastability threshold for two-dimensional bootstrap percolation. *Probab. Th. Rel. Fields* **125**, 195–224.

Horowitz, J. L. (2003). Bootstrap methods for Markov processes. *Econometrica* **71**, 1049–1082.

Huang, W.-M. (2003). On tests for proportion: A preliminary report. In *ASA Proceedings of the Joint Statistical Meetings, 1905*, American Statistical Association, Alexandria, VA.

Iglesias, P. M. C., and González-Manteiga, W. (2003). Bootstrap for the conditional distribution function with truncated and censored data. *Ann. Inst. Statist. Math.* **55**, 331–357.

Inoue, A., and Kilian, L. (2003). The continuity of the limit distribution in the parameter of interest is not essential for the validity of the bootstrap. *Econ. Theory* **19**, 944–961.

Janssen, A., and Pauls, T. (2003). How do bootstrap and permutation tests work? *Ann. Statist.* **31**, 768–806.

Jeng, S.-L. (2003). Inferences for the fatigue life model based on the Birnbaum–Saunders distribution. *Commun. Statist. Simul. Comput.* **32**, 43–60.

Jiménez-Gamero, M. D., Muñoz-García, J., and Pino-Mejías, R. (2003). Bootstrapping parameter estimated degenerate *U* and *V* statistics. *Statist. Probab. Lett.* **61**, 61–70.

Jing, B.-Y., Shao, Q.-M., and Wang, Q. (2003). Self-normalized Cramer-type large deviations for independent random variables. *Ann. Probab.* **31**, 2167–2215.

Karlis, D., and Xekalaki, E. (2003). Choosing initial values for the EM algorithm for finite mixtures. *Comput. Statist. Data Anal.* **41**, 577–590.

Kauermann, G., and Opsomer, J. D. (2003). Local likelihood estimation in generalized additive models. *Scand J. Statist.* **30**, 317–337.

Kaufman, S. (2003). The efficiency of the bootstrap under a locally random assumption for systematic samples. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 2097–2102. American Statistical Association, Alexandria, VA.

Kim, J. H. (2003). Forecasting autoregressive time series with bias-corrected parameter estimators. *Int. J. For.* **19**, 493–502.

Kim, Y., and Lee, J. (2003). Bayesian bootstrap for proportional hazards models. *Ann. Statist.* **31**, 1905–1922.

King, J. E. (2003). Bootstrapping confidence intervals for robust measures of association. *J. Mod. Appl. Statist. Meth.* **2**, 512–519.

Knight, K. (2003). On the second order behaviour of the bootstrap of *L1 regression estimators*. *J. Iranian Statist. Soc.* **2**, 21–42.

Kosorok, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Mult. Anal.* **84**, 299–318.

Kreiss, J.-P., and Paparoditis, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Ann. Statist.* **31**, 1923–1955.

Kuhnert, P. M., and Mengersen, K. (2003). Reliability measures for local nodes assessment in classification trees. *J. Comput. Graph. Statist.* **12**, 398–416.

Lahiri, S. N. (2003a). *Resampling Methods for Dependent Data*. Springer-Verlag, New York.*

Lahiri, S. N. (2003b). A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence. *Ann. Statist.* **31**, 613–641.

Lahiri, S. N. (2003c). *Validity of block bootstrap method for irregularly spaced spatial data under nonuniform stochastic designs*. Preprint, Department of Statistics, Iowa State University.

Lahiri, S. N. (2003d). Central limit theorems for weighted sums under some stochastic and fixed spatial sampling designs. *Sankhya A.* **65**, 356–388.

Lahiri, S. N., Furukawa, K., and Lee, Y.-D. (2003). *A nonparametric plug-in method for selecting optimal block design length*. Preprint, Department of Statistics, Iowa State University.

Lamarche, J.-F. (2003). A robust bootstrap test under heteroskedasticity. *Econ. Lett.* **79**, 353–359.

Langlet, É. R., Faucher, D., and Lesage, É. (2003). An application of the bootstrap variance estimation method to the Canadian participation and activity limitation survey. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 2299–2306. American Statistical Association, Alexandria, VA.

Lee, S. M. S., and Young, G. A. (2003). Prepivoting by weighted bootstrap iteration. *Biometrika* **90**, 393–410.

Li, Y., Ryan, L., Bellamy, S., and Satten, G. A. (2003). Inference on clustered survival data using imputed frailties. *J. Comput. Graph. Statist.* **12**, 640–662.

Li, Q., Hsiao, C., and Zinn, J. (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *J. Econ.* **112**, 295–325.

Liquet, B., Sakarovitch, C., and Commenges, D. (2003). Bootstrap choice of estimators in parametric and semiparametric families: An extension of EIC. *Biometrics* **59**, 172–178.

Lobato, I. N. (2003). Testing for nonlinear autoregression. *J. Bus. Econ Statist.* **21**, 164–173.

Malzahn, D., and Opper, M. (2003). Learning curves and bootstrap estimates for inference with Gaussian processes: A statistical mechanics study. *Complexity* **8**, 57–63.

Mannan, H. R., and Koval, J. J. (2003). Latent mixed Markov modelling of smoking transitions using Monte Carlo bootstrapping. *Statist. Meth. Med. Res.* **12**, 125–146.

Mazucheli, J., Louzada-Neto, F., and Achcar, J. A. (2003). Lifetime models with nonconstant shape parameters. *Revstat. Statist. J.* **1**, 25–39.

Moon, H., Ahn, H., and Kodell, R. L. (2003). Bootstrap adjustment of asymptotic normal tests for animal carcinogenicity data. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 2884–2888. American Statistical Association, Alexandria, VA.

Moore, D.S., McCabe, G.P., Duckworth, W.M., and Sclove, S.L. (2003). *Introduction to the Practice of Statistics*. 5th Edition (online). W.H. Freeman and Company, New York.*

Nelson, P. I., and Kemp, K. E. (2003). Testing for the presence of a maverick judge. *Commun. Statist. Th. Meth.* **32**, 807–826.

Orbe, J., Ferreira, E., and Núñez-Antón, V. (2003). Censored partial regression. *Biostatistics* **4**, 109–121.

Pandey, M. D., Gelder, P. H. A. J. M. Van, and Vrijling, J. K. (2003). Bootstrap simulations for evaluating the uncertainty associated with peaks-over-threshold estimates of extreme wind velocity. *EnvironMetrics* **14**, 27–43.

Paparoditis, E., and Politis, D. N. (2003). Residual-based block bootstrap for unit root testing. *Econometrica* **71**, 813–855.

Park, J. Y. (2003). Bootstrap unit root tests. *Econometrica* **71**, 1845–1895.

Pedres-Neto, P. R., Jackson, D. A., and Somers, K. M. (2003). Giving meaningful interpretation to ordination axes: Assessing loading significance in principal component analysis. *Ecology* **84**, 2347–2363.

Perakis, M., and Xekalaki, E. (2003). On a process capability index for asymmetric specifications. *Commun. Statist. Th. Meth.* **32**, 1459–1492.

Pigeot, I., Schäfer, J., Röhmel, J., and Hauschke, D. (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statist. Med.* **22**, 883–899.

Politis, D. N., and White, H. (2003). Automatic block-length selection for the dependent bootstrap. Preprint, Department of Mathematics, University of California at San Diego.

Politis, K. (2003). Semiparametric estimation for non-ruin probabilities. *Scand. Act. J.* **2003**, 75–96.

Poole, W. K., Gard, C. C., Das, A., and Bada, H. S. (2003). Sequential estimation of adjusted attributable risk in a cross-sectional study: A bootstrap approach. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 3335–3341. American Statistical Association, Alexandria, VA.

Psaradakis, Z. (2003). A bootstrap test for symmetry of dependent data based on a Kolmogorov–Smirnov type statistic. *Commun. Statist. Simul. Comput.* **32**, 113–126.

Qin, J., and Zhang, B. (2003). Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika* **90**, 585–596.

Raghunathan, T. E., Reiter, J. P., and Rubin, D. B. (2003). Multiple imputation for statistical disclosure limitation. *J. Official Statist.* **19**, 1–16.

Ren, J.-J. (2003). Goodness of fit tests with interval censored data. *Scand. J. Statist.* **30**, 211–226.

Robinson, J., Ronchetti, E., and Young, G. A. (2003). Saddlepoint approximations and tests based on multivariate *M*-estimates. *Ann. Statist.* **31**, 1154–1169.

Royston, P., and Sauerbrei, W. (2003). Stability of multivariable fractional polynomial models with selection of variables and transformations: A bootstrap investigation. *Statist. Med.* **22**, 639–659.

Samworth, R. (2003). A note on methods of restoring consistency to the bootstrap. *Biometrika* **90**, 985–990.

Sánchez, A., Ocaña, J., Utzet, F., and Serra, L. (2003). Comparison of Prevosti genetic distances. *J. Statist. Plann. Inf.* **109**, 43–65.

Schlattmann, P. (2003). Estimating the number of components in a finite mixture model: The special case of homogeneity. *Comput. Statist. Data Anal.* **41**, 441–451.

Schweder, T. (2003). Abundance estimation from multiple photo surveys: Confidence distributions and reduced likelihoods for Bowhead whales off Alaska. *Biometrics* **59**, 974–983.

Singh, K., and Xie, M. (2003). Bootlier-plot—Bootstrap based outlier detection plot. *Sankhy* , **65**, 532–559.

Sjöstedt-de Luna, S., and Young, A. (2003). The bootstrap and kriging prediction intervals. *Scand. J. Statist.* **30**, 175–192.

Solow, A. R., Stone, L., and Rozdilsky, I. (2003). A critical smoothing test for multiple equilibria. *Ecology* **84**, 1459–1463.

Stevens, R. J. (2003). Evaluation of methods for interval estimation of model outputs, with application to survival models. *J. Appl. Statist.* **30**, 967–981.

Sugahara, C. N. (2003). Bootstrap confidence intervals for the probability of correctly predicting the next state of a finite-state Markov chain. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 4128–4133. American Statistical Association, Alexandria, VA.

Sullivan, R., Timmermann, A., and White, H. (2003). Forecast evaluation with shared data sets. *Int. J. Forecast.* **19**, 217–227.

Swensen, A. R. (2003a). Bootstrapping unit root tests for integrated processes. *J. Time Series Anal.* **24**, 99–126.

Swensen, A. R. (2003b). A note on the power of bootstrap unit root tests. *Econ. Theory* **19**, 32–48.

Tajvidi, N. (2003). Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions. *Extremes* **6**, 111–123.

Tarek, J., and Dufour, J.-M. (2003). Finite sample simulation-based inference in vector autoregressive models. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 2032–2038, American Statistical Association, Alexandria, VA.

Tollenaar, N., and Mooijaart, A. (2003). Type I errors and power of the parametric bootstrap goodness-of-fit test: Full and limited information. *Brit. J. Math. Statist. Psych.* **56**, 271–288.

Toms, J. D., and Lesperance, M. L. (2003). Piecewise regression: A tool for identifying ecological thresholds. *Ecology* **84**, 2034–2041.

Trumbo, B. E., and Suess, E. A. (2003). Using simulation methods in statistics instruction: Evaluating estimators of variability. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 4292–4297. American Statistical Association, Alexandria, VA.

Tse, S.-K., and Xiang, L. (2003). Interval estimation for Weibull-distributed life data under Type II progressive censoring with random removals. *J. Biopharm. Statist.* **13**, 1–16.

Van den Noortgate, W., and Onghena, P. (2003). A parametric bootstrap version of Hedges' homogeneity test. *J. Mod. Appl. Statist. Meth.* **2**, 73–79.

Vinod, H. (2003). Constructive ensembles for time series analysis to avoid unit root testing. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 4372–4377. American Statistical Association, Alexandria, VA.

Vos, P. W., and Hudson, S. (2003). Simulation study of conditional, bootstrap, and t confidence intervals in linear regression. *Commun. Statist. Simul. Comput.* **32**, 697–715.

Wang, F., and Wall, M. M. (2003). Incorporating parameter uncertainty into prediction intervals for spatial data modeled via a parametric variogram. *J. Agr. Biol. Environ. Statist.* **8**, 296–309.

Ye, Z., and Weiss, R. E. (2003). Using the bootstrap to select one of a new class of dimension reduction methods. *J. Am. Statist. Assoc.* **98**, 968–979.

Young, G. A. (2003). Better bootstrapping by constrained prepivoting. *Metron* **61**, 227–242.

Yu, Q., and Wong, G. Y. C. (2003). Semi-parametric MLE in simple linear regression analysis with interval-censored data. *Commun. Statist. Simul. Comput.* **32**, 147–163.

Yuan, K.-H., and Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *Brit. J. Math. Statist. Psych.* **56**, 93–110.

Zhang, L.-C. (2003). Simultaneous estimation of the mean of a binary variable from a large number of small areas. *J. Off. Statist.* **19**, 253–263.

Zhu, L. (2003). Model checking of dimension-reduction type for regression. *Statist. Sin.* **13**, 283–296.

Zuehlke, T. W. (2003). Business cycle duration dependence reconsidered. *J. Bus. Econ. Statist.* **21**, 564–569.

2004

Alonso, A. M., Peña, D., and Romo, J. (2004). Introducing model uncertainty in time series bootstrap. *Statist. Sin.* **14**, 155–174.

Amado, C., and Pires, A. M. (2004). Robust bootstrap with non random weights based on the influence function. *Commun. Statist. Simul. Comput.* **33**, 377–396.

Andrews, D. W. K. (2004). The block-block bootstrap: Improved asymptotic refinements. *Econometrica* **72**, 673–700.

Austin, P. C., and Tu, J. V. (2004). Bootstrap methods for developing predictive models. *Am. Statist.* **58**, 131–137.

Babu, G. J. (2004). A note on the bootstrapped empirical process. *J. Statist. Plann. Inf.* **126**, 587–589.

Babu, G. J., and Rao, C. R. (2004). Goodness-of-fit tests when parameters are estimated. *Sankhyā* **66**, 63–74.

Baringhaus, L., and Franz, C. (2004). On a new multivariate two-sample test. *J. Mult. Anal.* **88**, 190–206.

Bee, M. (2004). Testing for redundancy in normal mixture analysis. *Commun. Statist. Simul. Comput.* **33**, 915–936.

Bilder, C. R., and Loughin, T. M. (2004). Testing for marginal independence between two categorical variables with multiple responses. *Biometrics* **60**, 241–248.

Bryan, J. (2004). Problems in gene clustering based on gene expression data. *J. Mult. Anal.* **90**, 44–66.

Capitanio, A., and Conti, P. L. (2004). A Bayesian nonparametric approach to the estimation of the adjustment coefficient, with applications to insurance and telecommunications. *Sankhyā* **66**, 75–108.

Chan, K. S., and Tong, H. (2004). Testing for multimodality with dependent data. *Biometrika* **91**, 113–123.

Chan, V., Lahiri, S. N., and Meeker, W. Q. (2004). Block bootstrap estimation of the distribution of cumulative outdoor degradation. *Technometrics* **46**, 215–224.

Chou, C.-Y., Chen, C.-H., and Liu, H.-R. (2004). Interval estimation for the smaller-the-better type of signal-to-noise ratio using bootstrap method. *Quality Eng.* **17**, 151–163.

Chung, H.-C., and Han, C.-P. (2004). Bootstrap confidence intervals for classification error rate in circular models when a block of observations is missing. In *ASA*

Proceedings of the Joint Statistical Meetings, pp. 2427–2431. American Statistical Association, Alexandria, VA.

Clarke, P. S., and Smith, P. W. F. (2004). Interval estimation for log-linear models with one variable subject to non-ignorable non-response. *J. R. Statist. Soc. B.* **66**, 357–368.

Corradi, V., and Swanson, N. R. (2004). Some recent developments in predictive accuracy testing with nested models and (generic) nonlinear alternatives. *J. Forecast.* **20**, 185–199.

Cribari-Neto, F., and Zarkos, S. (2004). Leverage-adjusted heteroskedastic bootstrap methods. *J. Statist. Comput. Simul.* **74**, 215–232.

Cruz, F. R. B., Colosimo, E. A., and Smith, J. M. (2004). Sample size corrections for the maximum partial likelihood estimator. *Commun. Statist. Simul. Comp.* **33**, 35–47.

Dai, M., and Guo, W. (2004). Multivariate spectral analysis using Cholesky decomposition, *Biometrika* **91**, 629–643.

Delaigle, A., and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.* **56**, 19–47.

Domínguez, M. A. (2004). On the power of bootstrapped specification tests. *Econ. Rev.* **23**, 215–228.

Derado, G., Mardia, K., Patrangenaru, V., and Thompson, H. (2004). A shape-based glaucoma index for tomographic images. *J. Appl. Statist.* **31**, 1241–1248.

Ekström, M., and Luna, S. S.-D. (2004). Subsampling methods to estimate the variance of sample means based on nonstationary spatial data with varying expected values. *J. Am. Statist. Assoc.* **99**, 82–95.

Fan, J., and Yim, T. H. (2004). A cross-validation method for estimating conditional densities. *Biometrika* **91**, 819–834.

Fledelius, P., Guillen, M., Nielsen, Jens P., and Vogelius, M. (2004). Two-dimensional hazard estimation for longevity analysis. *Scand. Act. J.* **2004**, 133–156.

Franco, G. C., and Reisen, V. A. (2004). Bootstrap techniques in semiparametric estimation methods for ARFIMA models: A comparison study. *Comput. Statist.* **19**, 243–259.

Franke, J., Neumann, M. H., and Stockis, J.-P. (2004). Bootstrapping nonparametric estimators of the volatility function. *J. Econ.* **118**, 189–218.

Fushiki, T., Komaki, F., and Aihara, K. (2004). On parametric bootstrapping and Bayesian prediction. *Scand. J. Statist.* **31**, 403–416.

Gelman, A. (2004). Exploratory data analysis for complex models. *J. Comput. Graph. Statist.* **13**, 755–779.

Ghosh, M., Zidek, J. V., Maiti, T., and White, R. (2004). The use of the weighted likelihood in the natural exponential families with quadratic variance. *Canad. J. Statist.* **32**, 139–157.

Gijbels, I., and Goderniaux, A.-C. (2004a). Bandwidth selection for changepoint estimation in nonparametric regression. *Technometrics* **46**, 76–86.

Gijbels, I., and Goderniaux, A.-C. (2004b). Data-driven discontinuity detection in derivatives of a regression function. *Commun. Statist. Th. Meth.* **33**, 851–871.

Godfrey, L. G., and Santos Silva, J. M. C. (2004). Bootstrap tests of nonnested hypotheses: Some further results. *Econ. Rev.* **23**, 325–340.

González-Manteiga, W., and Pérez-González, A. (2004). Nonparametric mean estimation with missing data. *Commun. Statist. Th. Meth.* **33**, 277–303.

Granger, C. W., Maasoumi, E., and Racine, J. (2004). A dependence metric for possibly nonlinear processes. *J. Time Series Anal.* **25**, 649–669.

Guo, H., and Krishnamoorthy, K. (2004). New approximate inferential methods for the reliability parameter in a stress–strength model: The normal case. *Commun. Statist. Th. Meth.* **33**, 1715–1731.

Hall, P., and Tajvidi, N. (2004). Prediction regions for bivariate extreme events. Aust. New Zealand *J. Statist.* **46**, 99–112.

Hall, P., and Wang, Q. (2004). Exact convergence rate and leading term in the Central Limit Theorem for Student's *t*-statistic. *Ann. Probab.* **32**, 1419–1437.

Härdle, W., Huet, S., Mammen, E., and Sperlich, S. (2004). Bootstrap inference in semiparametric generalized additive models. *Econ. Th.* **20**, 265–300.

Heffernan, J. E., and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B.* **66**, 497–530.

Hesterberg, T. C. (2004). Unbiasing the bootstrap: Bootknife sampling vs. smoothing. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 2924–2930. American Statistical Association, Alexandria.

Holgersson, H. E. T., and Shukur, G. (2004). Testing for multivariate heteroscedasticity. *J. Statist. Comput. Simul.* **74**, 879–896.

Holt, M., Stamey, J., Seaman J. R. J., and Young, D. (2004). A note on tests for interaction in quantal response data. *J. Statist. Comput. Simul.* **74**, 683–690.

Horvath, L., Kokoszka, P., and Teyssiere, G. (2004). Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *J. Statist. Comput. Simul.* **74**, 469–485.

Hutson, A. D. (2004). Exact nonparametric bootstrap confidence bands for the quantile function given censored data. *Commun. Statist. Simul. Comput.* **33**, 729–746.

Kaufman, S. (2004). Using the bootstrap in a two-stage sample design when some second-stage strata have only one unit allocated. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 3766–3773. American Statistical Association, Alexandria, VA.

Kim, J. H. (2004). Bootstrap prediction intervals for autoregression using asymptotically mean-unbiased estimators. *Int. J. Forecast.* **20**, 85–97.

Kim, T. Y., and Hwang, S. Y. (2004). Kernel matching scheme for block bootstrap of time series data. *J. Time Series Anal.* **25**, 199–216.

Kooperberg, C., and Stone, C. J. (2004). Comparison of parametric and bootstrap approaches to obtaining confidence intervals for logspline density estimation *J. Comput. Graph. Statist.* **13**, 106–122.

Landau, S., Ellison-Wright, I. C., and Bullmore, E. T. (2004). Tests for a difference in timing of physiological response between two brain regions measured by using functional magnetic resonance imaging. *J. R. Statist. Soc. C. Appl. Statist.* **53**, 63–82.

Lee, S.-M., Huang, L.-H. H., and Ou, S.-T. (2004). Band recovery model inference with heterogeneous survival rates. *Statist. Sin.* **14**, 513–531.

Len, L. F., and Tsai, C.-Ling (2004). Functional form diagnostics for Cox's proportional hazards model. *Biometrics* **60**, 75–84.

Li, Y., Lynch, C., Shimizu, I., and Kaufman, S. (2004). Imputation variance estimation by bootstrap method for the National Ambulatory Medical Care Survey. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 3883–3888. American Statistical Association, Alexandria.

Liang, H., Wang, S., Robins, J. M., and Carroll, R. J. (2004). Estimation in partially linear models with missing covariates. *J. Am. Statist. Assoc.* **99**, 357–367.

Liquet, B., and Commenges, D. (2004). Estimating the expectation of the log-likelihood with censored data for estimator selection. *Lifetime Data Anal.* **10**, 351–367.

Lo, K.-W. K., and Kelly, C. (2004). A bootstrap test for homogeneity of risk differences in a matched-pairs multi-center design. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 768–771. American Statistical Association, Alexandria, VA.

Loh, J. M., and Stein, M. L. (2004). Bootstrapping a spatial point process. *Statist. Sin.* **14**, 69–101.

MacNab, Y. C., Farrell, P. J., Gustafson, P., and Wen, S. (2004). Estimation in Bayesian disease mapping. *Biometrics* **60**, 865–873.

Montenegro, M., Colubi, A., Casals, M. R., and Gil, M. Á. (2004). Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika* **59**, 31–49.

Namba, A. (2004). Simulation studies on bootstrap empirical likelihood tests. *Commun. Statist. Simul. Comput.* **33**, 99–108.

Nordman, D., and Lahiri, S. N. (2004). On optimal spatial subsample size for variance estimation. *Ann. Statist.* **32**, 1981–2027.

Nordman, D., Lahiri, S. N., and Sibbertsen, P. (2004). *Empirical likelihood confidence intervals for the mean of a long range dependent process*. Preprint, Department of Statistics, Iowa State University.

Pascual, L., Romo, J., and Ruiz, E. (2004). Bootstrap predictive inference for ARIMA Processes. *J. Time Series Anal.* **25**, 449–465.

Pfeffermann, D., and Glickman, H. (2004). Mean square error approximation in small area estimation by use of parametric and nonparametric bootstrap. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 4167–4178, American Statistical Association, Alexandria, VA.

Pitarakis, J.-Y. (2004). Least squares estimation and tests of breaks in mean and variance under misspecification. *Econ. J. Online*, **7**, 32–54.

Pla, L. (2004). Bootstrap confidence intervals for the Shannon biodiversity index: A simulation study *J. Agr. Biol. Environ. Statist.* **9**, 42–56.

Politis, D. N., and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econ. Rev.* **23**, 53–70.

Presnell, B., and Boos, D. D. (2004). The ios test for model misspecification. *J. Am. Statist. Assoc.* **99**, 216–227.

Radulovic, D. (2004). Renewal type bootstrap for Markov chains. *Test* **13**, 147–192.

Rempala, G. A., and Szatzschneider, K. (2004). Bootstrapping parametric models of mortality. *Scand. Act. J.* **2004**, 53–78.

Reynolds, J. H., and Templin, W. D. (2004). Comparing mixture estimates by parametric bootstrapping likelihood ratios. *J. Agr. Biol. Environ. Statist.* **9**, 57–74.

Shen, X., Huang, H.-C., and Ye, J. (2004). Inference after model selection. *J. Am. Statist. Assoc.* **99**, 751–762.

Shi, Q., Zhu, Y., and Lu, J. (2004). Bootstrap approach for computing standard error of estimated coefficients in proportional odds model applied to correlated assessments in psychiatric clinical trial. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 845–854. American Statistical Association, Alexandria, VA.

Steinhorst, K., Wu, Y., Dennis, B., and Kline, P. (2004). Confidence intervals for fish out-migration estimates using stratified trap efficiency methods. *J. Agr. Biol. Environ. Statist.* **9**, 284–299.

Sverchkov, M., and Pfeffermann, D. (2004). Prediction of finite population totals based on the sample distribution. *Surv. Methodol.* **30**, 79–92.

Tamhane, A. C., and Logan, B. R. (2004).A superiority/equivalence approach to one-sided tests on multiple endpoints in clinical trials. *Biometrika* **91**, 715–727.

Taper, M. L. (2004). Bootstrapping dependent data in ecology. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 2994–2999. American Statistical Association, Alexandria, VA.

Troendle, J. F., Korn, E. L., and McShane, L. M. (2004). An example of slow convergence of the bootstrap in high dimensions. *Am. Statist.* **58**, 25–29.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., and Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *J. Math. Psych.* **48**, 28–50.

Willemain, T. R., Smart, C. N., and Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *Int. J. Forecast.* **20**, 375–387.

Wilson, M. D., McCormick, W. P., and Hinton, T. G. (2004). The maximally exposed individual? Comparison of maximum likelihood estimation of high quantiles to an extreme value estimate. *Risk Anal.* **24**, 1143–1151.

Xia, Y., Li, W. K., Tong, H., and Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statist. Sin.* **14**, 1–28.

Zhang, B. (2004). Assessing goodness-of-fit of categorical regression models based on case-control data. *Aust. New Zealand J. Statist.* **46**, 407–423.

Zhang, H. H., Wahba, G., Lin, Y., Voelker, M., Ferris, M., Klein, R., and Klein, B. (2004). Variable selection and model building via likelihood basis pursuit. *J. Am. Statist. Assoc.* **99**, 659–672.

Zheng, J., and Frey, H. C. (2004a). Quantification of variability and uncertainty using mixture distributions: Evaluation of sample size, mixing weights, and separation between components. *Risk Anal.* **24**, 553–571.

Zhao, Y., and Frey, H. C. (2004b). Quantification of variability and uncertainty for censored data sets and application to air toxic emission factors. *Risk Anal.* **24**, 1019–1034.

Zhu, J., and Morgan, G. D. (2004). Comparison of spatial variables over subregions using a block bootstrap. *J. Agr. Biol. Environ. Statist.* **9**, 91–104.

2005

Aldridge, G., and Bowman, D. (2005). Bayesian bootstrap methods for developmental toxicity studies. *J. Statist. Comput. Simul.* **75**, 81–91.

Baklizi, A. (2005). A continuously adaptive rank test for shift in location. *Aust. New Zealand J. Statist.* **47**, 203–209.

Babu, G. J. (2005). Bootstrap techniques for signal processing, by Abdelhak M. Zoubir and d. Robert iskander. *Technometrics* **47**, 374–375.

Banerjee, M., and Wellner, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32**, 405–424.

Bretz, F., Pinheiro, J., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose–response studies. *Biometrics* **61**, 738–748.*

Brouhns, N., Denuit, M., and Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scand. Act. J.* **32**, 212–224.

Chavez-Demoulin, V., and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *J. R. Statist. Soc. C, Appl. Statist.* **54**, 207–222.

Cheung, Y. K. (2005). Exact two-sample inference with missing data. *Biometrics* **61**, 524–531.

Chiang, C.-T., Wang, M.-C., and Huang, C.-Y. (2005). Kernel estimation of rate function for recurrent event data. *Scand. J. Statist.* **32**, 77–91.

Efron, B. (2005). Bayesians, frequentists, and scientists. *J. Am. Statist. Assoc.* **100**, 1–5.

Einmahl, J., and Rosalsky, A. (2005). General weak laws of large numbers for bootstrap sample means. *Stoch. Anal. Appl.* **23**, 853–869.

Feng, H., Willemain, T. R., and Shang, N. (2005). Wavelet-based bootstrap for time series analysis. *Commun. Statist. Simul. Comput.* **34**, 393–413.

Fernández de Castro, B., Guillas, S., and González-Manteiga, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics* **47**, 212–222.

Flachaire, E. (2005). More efficient tests robust to heteroskedasticity of unknown form. *Econ. Rev.* **24**, 219–241.

Fletcher, D., MacKenzie, D., and Villouta, E. (2005). Modelling skewed data with many zeros: A simple approach combining ordinary and logistic regression. *Environ. Ecol. Statist.* **12**, 45–54.

Freedman, D. A. (2005). *Statistical Models: Theory and Practice*. Cambridge University Press, Cambridge.

Galindo-Garre, F., and Vermunt, J. K. (2005). Testing log-linear models with inequality constraints: A comparison of asymptotic, bootstrap, and posterior predictive *P*-values. *Statist. Neerl.* **59**, 82–94.

Gonçalves, S., and White, H. (2005). Bootstrap standard error estimates for linear regression. *J. Am. Statist. Assoc.* **100**, 970–979.

Hall, P., and Samworth, R. J. (2005). Properties of bagged nearest neighbour classifiers. *J. R. Statist. Soc. B.* **67**, 363–379.

Harris, I. R., and Burch, B. D. (2005). Measuring relative importance of sources of variation without using variance. *Am. Statist.* **59**, 217–222.

Ho, Y. H. S., and Lee, S. M. S. (2005). Iterated smoothed bootstrap confidence intervals for population quantiles. *Ann. Statist.* **33**, 437–462.

Hothorn, T. L., Friedrich, Z. A., and Hornik, K. (2005). The design and analysis of benchmark experiments. *J. Comput. Graph. Statist.* **14**, 675–699.

Hui, T. Modarres, R., and Zheng, G. (2005). Bootstrap confidence interval estimation of mean via ranked set sampling linear regression. *J. Statist. Comput. Simul.* **75**, 543–553.

Jung, B., Cheol, J. M., and Lee, J. W. (2005). Bootstrap tests for overdispersion in a zero-inflated Poisson regression model. *Biometrics* **61**, 626–628.

Kocherginsky, M., He, X., and Mu, Y. (2005). Practical confidence intervals for regression quantiles. *J. Comput. Graph. Statist.* **14**, 41–55.

Kundu, D., and Gupta, R. D. (2005). Estimation of P[y < X] for generalized exponential distribution. *Metrika* **61**, 291–308.

Kuonen, D. (2005). Studentized bootstrap confidence intervals based on M-estimates. *J. Appl. Statist.* **32**, 443–460.

Lai, P. Y., and Lee, S. M. S. (2005). An overview of asymptotic properties of L_p regression under general classes of error distributions. *J. Am. Statist. Assoc.* **100**, 446–458.

Lahiri, S. N. (2005a). Consistency of the jackknife-after-bootstrap variance estimator for the bootstrap quantiles of a studentized statistic. *Ann. Statist.* **33**, 2475–2506.

Lahiri, S. N. (2005b). A note on the subsampling, method under long-range dependence. Preprint, Department of Statistics, Iowa State University.

Lahiri, S. N., and Zhu, J. (2005). Resampling methods for spatial regression models under a class of stochastic designs. Preprint, Department of Statistics, Iowa State University.

Lazar, N. A. (2005). Assessing the effect of individual data points on inference from empirical likelihood. *J. Comput. Graph. Statist.* **14**, 626–642.

Li, Y., and Williams, P. D. (2005). A new multiple-bootstrap-datasets presentation method for confidentiality protection. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 1306–1313. American Statistical Association, Alexandria, VA.

Machado, J. A. F., and Parente, P. (2005). Bootstrap estimation of covariance matrices via the percentile method. *Econ. J. Online* **8**, 70–78.

Marsh, L. C. (2005). Aspects of the exact finite sample distribution of the bootstrap. In *ASA Proceedings of the Joint Statistical Meetings*, pp. 906–913. American Statistical Association, Alexandria, VA.

Martin-Magniette, M. L. (2005). Nonparametric estimation of the hazard function by using a model selection method: Estimation of cancer deaths in Hiroshima atomic bomb survivors. *J. R. Statist. Soc. C. Appl. Statist.* **54**, 317–331.

Moon, H., Ahn, H., Lee, J. J., and Kodell, R. L. (2005). A weight-adjusted Peto's test when cause of death is not assigned. *Environ. Ecol. Statist.* **12**, 95–113.

Moore, D.S., and McCabe, G.P. (2005). *Introduction to the Practice of Statistics*. W.H. Freeman and Company, New York.*

Obenchain, R., Robinson, R., and Swindle, R. (2005). Cost-effectiveness inferences from bootstrap quadrant confidence levels: Three degrees of dominance. *J. Biopharm. Statist.* **15**, 419–436.

Paparoditis, E. (2005). Testing the fit of a vector autoregressive moving average model. *J. Time Series Anal.* **26**, 543–568.

Paparoditis, E., and Politis, D. N. (2005). Bootstrapping unit root tests for autoregressive time series. *J. Am. Statist. Assoc.* **100**, 545–553.

Perez, C. J., Martin, J., Rufo, M. J., and Rojano, C. (2005). Quasi-random sampling importance resampling. *Commun. Statist. Simul. Comput.* **34**, 97–111.

Raqab, M. Z., and Kundu, D. (2005). Comparison of different estimators of $P[y < X]$ for a scaled Burr type X distribution. *Commun. Statist. Simul. Comput.* **34**, 465–483.

Reiczigel, J., Zakariás, I., and Rózsa, L. (2005). A bootstrap test of stochastic equality of two populations. *Am. Statist.* **59**, 156–161.

Romano, J. P., and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Am. Statist. Assoc.* **100**, 94–108.

Samworth, R. (2005). Small confidence sets for the mean of a spherically symmetric distribution. *J. R. Statist. Soc. B* **67**, 343–361.

Şentürk, D., and Müller, H.-G. (2005). Covariate-adjusted regression. *Biometrika* **92**, 75–89.

Schlattmann, P. (2005). On bootstrapping the number of components in finite mixtures of Poisson distributions. *Statist. Comput.* **15**, 179–188.

Shoung, J.-M., Altan, S., and Cabrera, J. (2005). Double bootstrapping a tolerance limit. *J. Biopharm. Statist.* **15**, 367–373.

Shoukri, M. M., Chaudhary, M. A., and Al-Halees, A. (2005). Estimating $P(y < X)$ when X and y are paired exponential variables.*J. Statist. Comput. Simul.* **75**, 25–38.

Singh, K., Strawderman, W. E., and Xie, M. (2005). Combining information from independent sources through confidence distributions. *Ann. Statist.* **33**, 159–183.

Tubert-Bitter, P., Letierce, A., Bloch, D. A., and Kramar, A. (2005). A nonparametric comparison of the effectiveness of treatments: A multivariate toxicity-penalized approach. *J. Biopharm. Statist.* **15**, 129–142.

Turk, P., and Borkowski, J. J. (2005). A review of adaptive cluster sampling: 1990–2003. *Environ. Ecol. Statist.* **12**, 55–94.

Ugarte, M. D., Ibáñez, B., and Militino, A. F. (2005). Detection of spatial variation in risk when using car models for smoothing relative risks. *Stoch. Environ. Res. Risk Assess.* **19**, 33–40.

Yin, G., and Cai, J. (2005). Quantile regression models with multivariate failure time data. *Biometrics* **61**, 151–161.

Zheng, J., and Frey, H. C. (2005). Quantitative analysis of variability and uncertainty with known measurement error: Methodology and case study. *Risk Anal.* **25**, 663–675.

Zhou, X. H. (2005). Nonparametric confidence intervals for the one- and two-sample problems. *Biostatistics*, **6**, 187–200.

2006

Baklizi, A. (2006). Asymptotic and resampling-based confidence intervals for $P(X < y)$. *Commun. Statist. Simul. Comput.* **35**, 295–307.

Balogh, J., and Bollobás, B. (2006). Bootstrap percolation on the hypercube. *Probab. Th. Rel. Fields* **134**, 624–648.

Bertail, P., and Tressou, J. (2006). Incomplete generalized U-statistics for food risk assessment. *Biometrics* **62**, 66–74.

Bühlmann, P., and Lutz, R. W. (2006). Boosting Algorithms: with Application to Bootstrapping Multivariate Time Series. In *Fromtiers in Statistics* (J. Fan and H. L. Koul editors), pp. 209–230. Imperial College Press, London.*

Cabras, S., Mostallino, G., and Racugno, W. (2006). A nonparametric bootstrap test for the equality of coefficients of variation. *Commun. Statist. Simul. Comput.* **35**, 715–726.

Cadarso-Suárez, C., Roca-Pardiñas, J., Molenberghs, G., Faes, C., Nácher, V., Ojeda, S., and Acuña, C. (2006). Flexible modelling of neuron firing rates across different experimental conditions: An application to neural activity in the prefrontal cortex during a discrimination task. *J. R. Statist. Soc. C.* **55**, 431–447.

Cavaliere, G., and Taylor, A. M. R. (2006). Testing the null of Co-integration in the presence of variance breaks. *J. Time Series Anal.* **27**, 613–636.

Chakraborti, S., Hong, B., and van de Wiel, M. A. (2006). A note on sample size determination for a nonparametric test of location. *Technometrics* **48**, 88–94.

Chan, K. Y. F., Lee, S. M. S., and Ng, K. W. (2006). Minimum variance unbiased estimation based on bootstrap iterations. *Statist. Comput.* **16**, 267–277.

Chen, M., Kianifard, F., and Dhar, S. (2006). A bootstrap-based test for establishing noninferiority in clinical trials. *J. Bio. Statist.* **16**, 357–363.

Cheol J. B., Jhun, M., and Heun S., S. (2006). Testing for overdispersion in a censored Poisson regression model. *Statistics* **40**, 533–543.

Choudhary, P. K., and Tony Ng, H. K. (2006). Assessment of agreement under nonstandard conditions using regression models for mean and variance. *Biometrics* **62**, 288–296.

Colubi, A., Santos Dominguez-Menchero, J., and Gonzalez-Rodriguez, G. (2006). Testing constancy for isotonic regressions. *Scand. J. Statist.* **33**, 463–475.

Cox, D. R. (2006). *Principles of Statistical Inference*. Cambridge University Press, Cambridge.*

de Silva, B., and Waikar, V. (2006). A sequential approach to the Behrens–Fisher problem. *Seq. Anal.* **25**, 311–326.

Dette, H., Podolskij, M., and Vetter, M. (2006). Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing. *Scand. J. Statist.* **33**, 259–278.

DiCiccio, T. J., Monti, A. C., and Young, G. A. (2006). Variance stabilization for a scalar parameter. *J. R. Statist. Soc. B.* **68**, 281–303.

Dikta, G., Kvesic, M., and Schmidt, C. (2006). Bootstrap approximations in model checks for binary data. *J. Am. Statist. Assoc.* **101**, 521–530.

Droge, B. (2006). Book Review of *Permutation, Parametric, and Bootstrap Tests of Hypotheses* (by Philip Good). *Metrika* **64**, 249–250.

Escanciano, J. C. (2006). Goodness-of-fit tests for linear and nonlinear time series models. *J. Am. Statist. Assoc.* **101**, 531–541.

Fan, J., and Koul, H. L. (2006). *Frontiers in Statistics*. Imperial Collage Press, London.*

Franco, G., Reisen, V., and Barros, P. (2006). Unit root tests using semi-parametric estimators of the long-memory parameter. *J. Statist. Comput. Simul.* **76**, 727–735.

Guarte, J., and Barrios, E (2006). Estimation under purposive sampling. *Commun. Statist. Simul. Comput.* **35**, 277–284.

Godfrey, L. G., Orme, C. D., and Santos Silva, J. M. C. (2006). Simulation-based tests for heteroskedasticity in linear regression models: Some further results. *Econ. J. Online* **9**, 76–97.

Hall, P., and Maiti, T. (2006a). On parametric bootstrap methods for small area prediction. *J. R. Statist. Soc. B.* **68**, 221–238.

Hall, P., and Maiti, T. (2006b). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *Ann. Statist.* **34**, 1733–1750.

Hall, P., and Vial, C. (2006a). Assessing extrema of empirical principal component functions. *Ann. Statist.* **34**, 1518–1544.

Hall, P., and Vial, C. (2006b). Assessing the finite dimensionality of functional data. *J. R. Statist. Soc. B.* **68**, 689–705.

He, Y., and Raghunathan, T. E. (2006). Tukey's *gh* distribution for multiple imputation. *Am. Statist.* **60**, 251–256.

Hodoshima, J., and ando, M. (2006). The effect of non independence of explanatory variables and error term and heteroskedasticity in stochastic regression models. *Commun. Statist. Simul. Comput.* **35**, 361–405.

Hoff, A. (2006). Bootstrapping Malmquist indices for Danish seiners in the North sea and skagerrak1. *J. Appl. Statist.* **33**, 891–907.

Hu, T.-C., Cabrera, M., and Volodin, A. (2006). Almost sure lim Sup behavior of dependent bootstrap means. *Stoch. Anal. Appl.* **24**, 939–942.

Inoue, A. (2006). A bootstrap approach to moment selection. *Econ. J. Online* **9**, 48–75.

Jeske, D. R., and Chakravartty, A. (2006). Effectiveness of bootstrap bias correction in the context of clock offset estimators. *Technometrics* **48**, 530–538.

Kwon, H.-H., and Moon, Y.-I. (2006). Improvement of overtopping risk evaluations using probabilistic concepts for existing dams. *Stoch. Environ. Res. Risk Assess.* **20**, 223–237.

Lahiri, S. N. (2006). Bootstrap Methods: A Review. In *Frontiers in Statistics* (J. Fan and H. L. Koul, editors), pp. 231–265. Imperial College Press, London.*

Lahiri, S. N., and Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Ann. Statist.* **34**, 1774–1813.

Lee, S. M. S., and Pun, M. C. (2006). On m out of n bootstrapping for nonstandard *M*-estimation with nuisance parameters. *J. Am. Statist. Assoc.* **101**, 1185–1197.

Levina, E., and Bickel, P. J. (2006). Texture synthesis and nonparametric resampling of random fields. *Ann. Statist.* **34**, 1751–1773.

Li, Y., and Ryan, L. (2006). Inference on survival data with covariate measurement error—an imputation-based approach. *Scand. J. Statist.* **33**, 169–190.

Martin, M., and Roberts, S. (2006). An evaluation of bootstrap methods for outlier detection in least squares regression. *J. Appl. Statist.* **33**, 703–720.

Massonnet, G., Burzykowski, T., and Janssen, P. (2006). Resampling plans for frailty models. *Commun. Statist. Simul. Comput.* **35**, 497–514.

Neumeyer, N., Dette, H., and Nagel, E.-R. (2006). Bootstrap tests for the error distribution in linear and nonparametric regression models Aust. *New Zealand J. Statist.* **48**, 129–156.

Neumeyer, N., and Sperlich, S. (2006). Comparison of separable components in different samples. *Scand. J. Statist.* **33**, 477–501.

Ogden, R. T., and Tarpey, T. (2006). Estimation in regression models with externally estimated parameters. *Biostatistics* **7**, 115–129.

Omtzigt, P., and Fachin, S. (2006). The size and power of bootstrap and Bartlett-corrected tests of hypotheses on the cointegrating vectors. *Econ. Rev.* **25**, 41–60.

Pardo-Fernandez, J. C., and van Keilegom, I. (2006). Comparison of regression curves with censored responses. *Scand. J. Statist.* **33**, 409–434.

Park, Y., Choi, J. W., and Kim, H.-Y. (2006). Forecasting cause-age specific mortality using two random processes. *J. Am. Statist. Assoc.* **101**, 472–483.

Patterson, S., and Jones, B. (2006). *Bioequivalence and Statistics in Clinical Pharmacolgy*, Chapman & Hall/CRC, Boca Raton.*

Percival, D. B., and Constantine, W. L. B. (2006). Exact simulation of Gaussian time series from nonparametric spectral estimates with application to bootstrapping. *Statist. Comput.* **16**, 25–35.

Perez, T., and Pontius, J. (2006). Conventional bootstrap and normal confidence interval estimation under adaptive cluster sampling. *J. Statist. Comput. Simul.* **76**, 755–764.

Qin, G., and Zhou, X.-H. (2006). Empirical likelihood inference for the area under the ROC curve. *Biometrics* **62**, 613–622.

Saavedra, P., Santana, A., and Quintana, M. (2006). Pivotal quantities based on sequential data: A bootstrap approach. *Commun. Statist. Simul. Comput.* **35**, 1005–1018.

Salibián-Barrera, M., Van Aelst, S., and Willems, G. (2006). Principal components analysis based on multivariate *m*-estimators with fast and robust bootstrap. *J. Am. Statist. Assoc.* **101**, 1198–1211.

Sherman, M., Apanasovich, T., and Carroll, R. (2006). On estimation in binary autologistic spatial models. *J. Statist. Comput. Simul.* **76**, 167–179.

Sergeant, J. C., and Firth, D. (2006). Relative index of inequality: Definition, estimation, and inference. *Biostatistics* **7**, 213–224.

Tanizaki, H., Hamori, S., and Matsubayashi, Y. (2006). On least-squares bias in the AR(p) models: Bias correction using the bootstrap methods. *Statist. Papers* **47**, 109–124.

Temime, L., and Thomas, G. (2006). Estimation of balanced simultaneous confidence sets for SIR models. *Commun. Statist. Simul. Comput.* **35**, 803–812.

Townsend, R. L., Skalski, J. R., Dillingham, P., and Steig, T. W. (2006). Correcting bias in survival estimation resulting from tag failure in acoustic and radiotelemetry studies. *J. Agr. Biol. Environ. Statist.* **11**, 183–196.

Tressou, J. (2006). Nonparametric modeling of the left censorship of analytical data in food risk assessment. *J. Am. Statist. Assoc.* **101**, 1377–1386.

van der Laan, M., and Hubbard, A. (2006). Quantile-function based null distribution in resampling based multiple testing *Statist. Appl. Gen. Mol. Biol.* **5**,

Wang, J. (2006). Quadratic artificial likelihood functions using estimating functions. *Scand. J. Statist.* **33**, 379–390.

Wilcox, R., and Keselman, H. (2006). Detecting heteroscedasticity in a simple regression model via quantile regression slopes. *J. Statist. Comput. Simul.* **76**, 705–712.

Wolfsegger, M. J., and Jaki, T. (2006). Simultaneous confidence intervals by iteratively adjusted alpha for relative effects in the one-way layout. *Statist. Comput.* **16**, 15–23.

Zhu, Y., and Zeng, P. (2006). Fourier methods for estimating the central subspace and the central mean subspace in regression. *J. Am. Statist. Assoc.* **101**, 1638–1651.

2007

Dmitrienko, A., Chuang-stein, C., and D'Agostino, R. (editors) (2007). *Pharmaceutical Statistics Using SAS®: A Practical Guide*. SAS Institute, Inc., Cary, NC.*

Dodd, L. E., and Korn, E. L. (2007). The Bootstrap Variance of the Square of a Sample Mean. *Am. Statist.* **61**, 127–131.*

Gill, P. (2007). Efficient calculation of *P*-values in linear-statistic permutation significance tests. *J. Statist. Comput. Simul.* **77**, 55–61.

Klingenberg, B. (2007). A unified framework for the proof of concept and dose estimation with binary responses. Unpublished manuscript submitted to *Biometrics*.*

Zhu, J., and Lahiri, S. N. (2007). Bootstrapping the empirical distribution function of a spatial process. *Statist. Inf. Stoch. Proc.* **10**, 107–145.

Author Index

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition, by Michael R. Chernick

Subject Index

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition, by Michael R. Chernick