

Statistics in the Health Sciences

Theory, Applications, and Computing

Albert Vexler

Alan Hutson

This book should be appropriate for use both as a text and as a reference. This book delivers a "ready-to-go" well-structured product to be employed in developing advanced courses. In this book the readers can find classical and new theoretical methods, open problems and new procedures.

KEY FEATURES

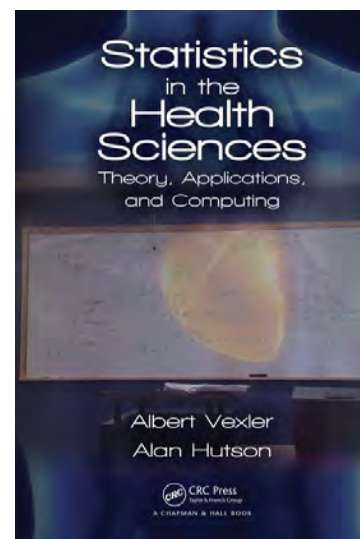
- A systematic compendium of intermediate and advanced methods in biostatistics
- A good balance of rigor in methodology and applications
- Software code based on both the R and SAS software packages to exemplify the presented topics.
- Detailed worked examples, exercises and exam questions, as well as student projects

SELECTED CONTENTS

Prelude: Preliminary Tools and Foundations
Characteristic Function Based Inference
Likelihood Tenet
Martingale Type Statistics and Their Applications
Bayes Factor
A Brief Review of Sequential Methods
A Brief Review of Receiver Operating Characteristic Curve Analyses
The Ville and Wald Inequality: Extensions and Applications
Brief Comments on Confidence Intervals and P-Values
Empirical Likelihood
Jackknife and Bootstrap Methods
Examples of Homework Questions
Examples of Exams
Examples of Courses Projects

SAVE 20% when you order online and enter Promo Code **FMQ13**

FREE standard shipping when you order online.



Catalog no. K31208

February 2018, 398 pp.

ISBN: 978-1-1381-9689-6

\$99.95 / £77.00

www.crcpress.com

e-mail: orders@crcpress.com

1-800-634-7064 • 1-561-994-0555 • +44 (0) 1235 400 524



CRC Press
Taylor & Francis Group

Statistics

"This very informative book introduces classical and novel statistical methods that can be used by theoretical and applied biostatisticians to develop efficient solutions for real-world problems encountered in clinical trials and epidemiological studies. The authors provide a detailed discussion of methodological and applied issues in parametric, semiparametric, and nonparametric approaches, including computationally extensive data-driven techniques, such as empirical likelihood, sequential procedures, and bootstrap methods. Many of these techniques are implemented using popular software such as R and SAS."

— Vlad Dragalin, Professor, Johnson & Johnson
Spring House, Pennsylvania

"It is always a pleasure to come across a new book that covers nearly all facets of a branch of science one thought was so broad, so diverse, and so dynamic that no single book could possibly hope to capture all of the fundamentals as well as directions of the field. The topics within the book's purview—fundamentals of measure-theoretic probability; parametric and nonparametric statistical inference; central limit theorems; basics of martingale theory; Monte Carlo methods; sequential analysis; sequential change-point detection—are all covered with inspiring clarity and precision. The authors are also very thorough and avail themselves of the most recent scholarship. They provide a detailed account of the state of the art, and bring together results that were previously scattered across disparate disciplines. This makes the book more than just a textbook: it is a panoramic companion to the field of Biostatistics. The book is self-contained, and the concise but careful exposition of material makes it accessible to a wide audience. This is appealing to graduate students interested in getting into the field, and also to professors looking to design a course on the subject."

— Aleksey S. Polunchenko, Department of Mathematical Sciences,
State University of New York at Binghamton

This book should be appropriate for use both as a text and as a reference. This book delivers a "ready-to-go" well-structured product to be employed in developing advanced courses.

In this book the readers can find classical and new theoretical methods, open problems, and new procedures.

The book presents biostatistical results that are novel to the current set of books on the market and results that are even new with respect to the modern scientific literature. Several of these results can be found only in this book.

- A systematic compendium of intermediate and advanced methods in biostatistics
- A good balance of rigor in methodology and applications
- Software code based on both the R and SAS software packages to exemplify the presented topics
- Detailed worked examples, exercises, and exam questions, as well as student projects



CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK



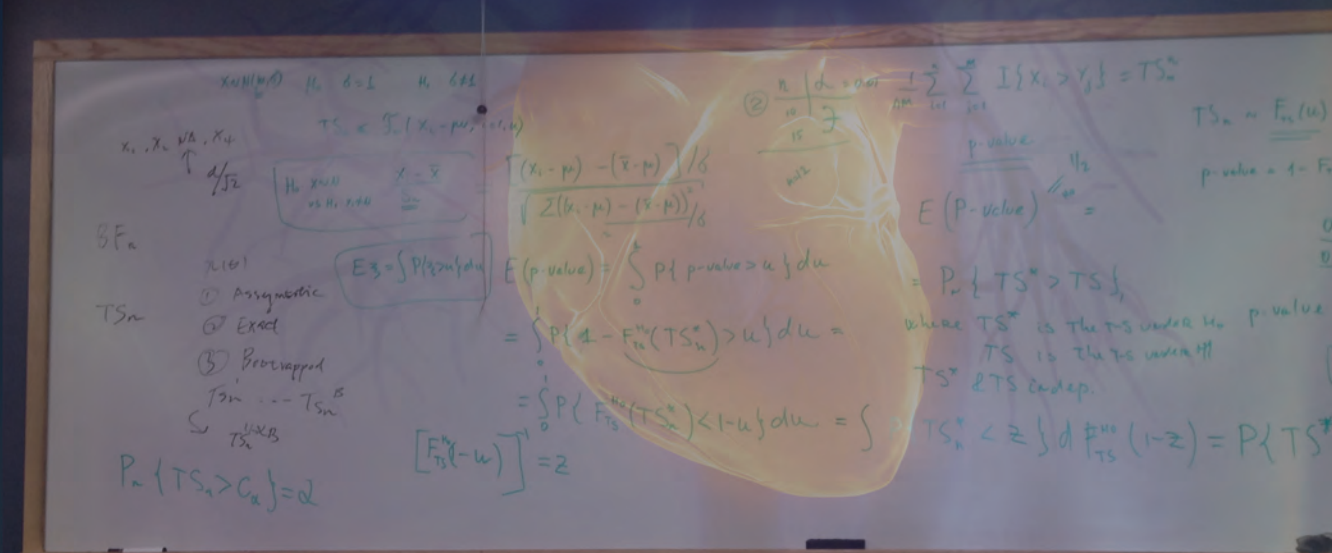
Statistics in the Health Sciences

Vexler
Hutson



Statistics in the Health Sciences

Theory, Applications
and Computing



Albert Vexler
Alan Hutson



CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK



Statistics in the Health Sciences

Theory, Applications
and Computing



Statistics in the Health Sciences

Theory, Applications and Computing

By
Albert Vexler
The State University of New York at Buffalo, USA
Alan D. Hutson
Roswell Park Cancer Institute, USA



CRC Press

Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2018 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper

International Standard Book Number-13: 978-1-138-19689-6 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Names: Vexler, Albert, author. | Hutson, Alan D., author.
Title: Statistics in the health sciences : theory, applications, and computing / Albert Vexler and Alan D. Hutson.
Description: Boca Raton, Florida : CRC Press, [2018] | Includes bibliographical references and index.
Identifiers: LCCN 2017031841 | ISBN 9781138196896 (hardback) | ISBN 9781315293776 (e-book) | ISBN 9781315293769 (e-book) | ISBN 9781315293752 (e-book) | ISBN 9781315293745 (e-book)
Subjects: LCSH: Medical statistics. | Medicine—Research—Statistical methods. | Epidemiology—Statistical methods. | Statistical hypothesis testing.
Classification: LCC R853.S7 V48 2018 | DDC 610.2/1—dc23
LC record available at <https://lcn.loc.gov/2017031841>

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>
and the CRC Press Web site at
<http://www.crcpress.com>

To my parents, Octyabrina and Alexander, and my son, David

Albert Vexler

To my wife, Brenda, and three kids, Nick, Chance, and Trey

Alan D. Hutson



Table of Contents

Preface.....	xiii
Authors	xvii

1. Prelude: Preliminary Tools and Foundations..... 1

1.1	Introduction	1
1.2	Limits	1
1.3	Random Variables	2
1.4	Probability Distributions	3
1.5	Commonly Used Parametric Distribution Functions.....	5
1.6	Expectation and Integration	6
1.7	Basic Modes of Convergence of a Sequence of Random Variables	8
1.7.1	Convergence in Probability	9
1.7.2	Almost Sure Convergence	9
1.7.3	Convergence in Distribution	9
1.7.4	Convergence in r th Mean	10
1.7.5	$O(\cdot)$ and $o(\cdot)$ Revised under Stochastic Regimes	10
1.7.6	Basic Associations between the Modes of Convergence	10
1.8	Indicator Functions and Their Bounds as Applied to Simple Proofs of Propositions	10
1.9	Taylor's Theorem	12
1.10	Complex Variables	14
1.11	Statistical Software: R and SAS.....	16
1.11.1	R Software.....	16
1.11.2	SAS Software	22

2. Characteristic Function–Based Inference..... 31

2.1	Introduction	31
2.2	Elementary Properties of Characteristic Functions	32
2.3	One-to-One Mapping	34
2.3.1	Proof of the Inversion Theorem	35
2.4	Applications	40
2.4.1	Expected Lengths of Random Stopping Times and Renewal Functions in Light of Tauberian Theorems.....	40
2.4.2	Risk-Efficient Estimation.....	46
2.4.3	Khinchin's (or Hinchin's) Form of the Law of Large Numbers.....	48
2.4.4	Analytical Forms of Distribution Functions.....	49

2.4.5	Central Limit Theorem	50
2.4.5.1	Principles of Monte Carlo Simulations	51
2.4.5.2	That Is the Question: Do Convergence Rates Matter?	54
2.4.6	Problems of Reconstructing the General Distribution Based on the Distribution of Some Statistics	54
2.4.7	Extensions and Estimations of Families of Distribution Functions	56
3.	Likelihood Tenet.....	59
3.1	Introduction	59
3.2	Why Likelihood? An Intuitive Point of View	61
3.3	Maximum Likelihood Estimation	63
3.4	The Likelihood Ratio	69
3.5	The Intrinsic Relationship between the Likelihood Ratio Test Statistic and the Likelihood Ratio of Test Statistics: One More Reason to Use Likelihood.....	73
3.6	Maximum Likelihood Ratio	76
3.7	An Example of Correct Model-Based Likelihood Formation.....	80
4.	Martingale Type Statistics and Their Applications	83
4.1	Introduction	83
4.2	Terminology	84
4.3	The Optional Stopping Theorem and Its Corollaries: Wald's Lemma and Doob's Inequality.....	90
4.4	Applications	93
4.4.1	The Martingale Principle for Testing Statistical Hypotheses	93
4.4.1.1	Maximum Likelihood Ratio in Light of the Martingale Concept	96
4.4.1.2	Likelihood Ratios Based on Representative Values.....	97
4.4.2	Guaranteed Type I Error Rate Control of the Likelihood Ratio Tests.....	100
4.4.3	Retrospective Change Point Detection Policies.....	100
4.4.3.1	The Cumulative Sum (CUSUM) Technique	101
4.4.3.2	The Shirayev–Roberts (SR) Statistic-Based Techniques.....	102
4.4.4	Adaptive (Nonanticipating) Maximum Likelihood Estimation	105
4.4.5	Sequential Change Point Detection Policies	109
4.5	Transformation of Change Point Detection Methods into a Shirayev–Roberts Form.....	113
4.5.1	Motivation.....	114
4.5.2	The Method.....	115

4.5.3	CUSUM versus Shirayev–Roberts.....	120
4.5.4	A Nonparametric Example.....	123
4.5.5	Monte Carlo Simulation Study.....	124
5.	Bayes Factor.....	129
5.1	Introduction	129
5.2	Integrated Most Powerful Tests	132
5.3	Bayes Factor.....	136
5.3.1	The Computation of Bayes Factors	137
5.3.1.1	Asymptotic Approximations.....	142
5.3.1.2	Simple Monte Carlo, Importance Sampling, and Gaussian Quadrature.....	147
5.3.1.3	Generating Samples from the Posterior	147
5.3.1.4	Combining Simulation and Asymptotic Approximations.....	148
5.3.2	The Choice of Prior Probability Distributions.....	149
5.3.3	Decision-Making Rules Based on the Bayes Factor	153
5.3.4	A Data Example: Application of the Bayes Factor	158
5.4	Remarks.....	161
6.	A Brief Review of Sequential Methods	165
6.1	Introduction	165
6.2	Two-Stage Designs.....	167
6.3	Sequential Probability Ratio Test.....	169
6.4	The Central Limit Theorem for a Stopping Time.....	177
6.5	Group Sequential Tests.....	179
6.6	Adaptive Sequential Designs	181
6.7	Futility Analysis.....	182
6.8	Post-Sequential Analysis.....	182
7.	A Brief Review of Receiver Operating Characteristic Curve Analyses.....	185
7.1	Introduction	185
7.2	ROC Curve Inference	186
7.3	Area under the ROC Curve	189
7.3.1	Parametric Approach	190
7.3.2	Nonparametric Approach.....	192
7.4	ROC Curve Analysis and Logistic Regression: Comparison and Overestimation	193
7.4.1	Retrospective and Prospective ROC	195
7.4.2	Expected Bias of the ROC Curve and Overestimation of the AUC	196
7.4.3	Example.....	198

7.5	Best Combinations Based on Values of Multiple Biomarkers	201
7.5.1	Parametric Method	202
7.5.2	Nonparametric Method	202
7.6	Remarks	202
8.	The Ville and Wald Inequality: Extensions and Applications	205
8.1	Introduction	205
8.2	The Ville and Wald inequality	208
8.3	The Ville and Wald Inequality Modified in Terms of Sums of iid Observations	209
8.4	Applications to Statistical Procedures	213
8.4.1	Confidence Sequences and Tests with Uniformly Small Error Probability for the Mean of a Normal Distribution with Known Variance	213
8.4.2	Confidence Sequences for the Median	215
8.4.3	Test with Power One	217
9.	Brief Comments on Confidence Intervals and p-Values	219
9.1	Confidence Intervals	219
9.2	p -Values	223
9.2.1	The EPV in the Context of an ROC Curve Analysis	226
9.2.2	Student's t -Test versus Welch's t -Test	227
9.2.3	The Connection between EPV and Power	229
9.2.4	t -Tests versus the Wilcoxon Rank-Sum Test	231
9.2.5	Discussion	236
10.	Empirical Likelihood	239
10.1	Introduction	239
10.2	Empirical Likelihood Methods	239
10.3	Techniques for Analyzing Empirical Likelihoods	242
10.3.1	Practical Theoretical Tools for Analyzing ELs	249
10.4	Combining Likelihoods to Construct Composite Tests and to Incorporate the Maximum Data-Driven Information	253
10.5	Bayesians and Empirical Likelihood	254
10.6	Three Key Arguments That Support the Empirical Likelihood Methodology as a Practical Statistical Analysis Tool	255
	Appendix	256
11.	Jackknife and Bootstrap Methods	259
11.1	Introduction	259
11.2	Jackknife Bias Estimation	261
11.3	Jackknife Variance Estimation	263

11.4	Confidence Interval Definition	263
11.5	Approximate Confidence Intervals	264
11.6	Variance Stabilization.....	265
11.7	Bootstrap Methods.....	266
11.8	Nonparametric Simulation	267
11.9	Resampling Algorithms with SAS and R.....	270
11.10	Bootstrap Confidence Intervals.....	279
11.11	Use of the Edgeworth Expansion to Illustrate the Accuracy of Bootstrap Intervals.....	284
11.12	Bootstrap- t Percentile Intervals.....	288
11.13	Further Refinement of Bootstrap Confidence Intervals	291
11.14	Bootstrap Tilting.....	303
12	Examples of Homework Questions.....	305
12.1	Homework 1.....	305
12.2	Homework 2.....	306
12.3	Homework 3.....	306
12.4	Homework 4.....	307
12.5	Homework 5.....	308
12.6	Homeworks 6 and 7	311
13	Examples of Exams.....	315
13.1	Midterm Exams.....	315
Example 1	315
Example 2	316
Example 3	317
Example 4	318
Example 5	319
Example 6	320
Example 7	322
Example 8	324
Example 9	325
13.2	Final Exams.....	326
Example 1	326
Example 2	328
Example 3	329
Example 4	331
Example 5	332
Example 6	333
Example 7	335
Example 8	336
Example 9	339
13.3	Qualifying Exams.....	340
Example 1	340
Example 2	342

Example 3.....	345
Example 4.....	348
Example 5.....	350
Example 6.....	354
Example 7.....	357
Example 8.....	359
14 Examples of Course Projects	363
14.1 Change Point Problems in the Model of Logistic Regression Subject to Measurement Errors.....	363
14.2 Bayesian Inference for Best Combinations Based on Values of Multiple Biomarkers.....	363
14.3 Empirical Bayesian Inference for Best Combinations Based on Values of Multiple Biomarkers.....	364
14.4 Best Combinations Based on Log Normally Distributed Values of Multiple Biomarkers	364
14.5 Empirical Likelihood Ratio Tests for Parameters of Linear Regressions.....	365
14.6 Penalized Empirical Likelihood Estimation	366
14.7 An Improvement of the AUC-Based Interference	366
14.8 Composite Estimation of the Mean based on Log-Normally Distributed Observations	366
References	369
Author Index	381
Subject Index	387

Preface—Please Read!



In all likelihood, the Universe of Statistical Science is “relatively” or “privately” infinite and expanding in a similar manner to our Universe, satisfying the property of a science that is alive. Hence, it would be absolute impossibility to include everything in one book written by humans. One of our major goals is to provide the readers a small but efficient “probe” that could assist in Statistical Space discoveries.

The primary objective of this book is to provide a compendium of statistical techniques ranging from classical methods through bootstrap strategies to modern recently developed statistical techniques. These methodologies may be applied to various problems encountered in health-related studies.

Historically, initial developments in statistical science were induced by real-life problems, when appropriate statistical instruments employed empirical arguments. Perhaps, since the eighteenth century, the heavy convolution between mathematics, probability theory, and statistical methods has provided the fundamental structures for correct statistical and biostatistical techniques. However, we cannot ignore a recent trend toward redundant simplifications of statistical considerations via so called “intuitive” and “applied” claims in complex statistical applications. It is our experience that applied statisticians or users often neglect the underlying postulates when implementing formal statistical procedures and with respect to the interpretation of their results. A very important motivation towards writing this book was to better refocus the scientist towards understanding the underpinnings of appropriate statistical inference in a well-rounded fashion. Maybe now is the time to draw more attention of theoretical and applied researchers to methodological standards in statistics and biostatistics? In contrast with

many biostatistical books, we focus on rigorous formal proof schemes and their extensions regarding different statistical principles. We also show the basic ingredients and methods for constructing and examining correct and powerful statistical processes.

The material in the book should be appropriate for use both as a text and as a reference. In our book readers can find classical and new theoretical methods, open problems, and new procedures across a variety of topics for their scholarly investigations. We present results that are novel to the current set of books on the market and results that are even new with respect to the modern scientific literature. Our aim is to draw the attention of theoretical statisticians and practitioners in epidemiology and/or clinical research to the necessity of new developments, extensions and investigations related to statistical methods and their applications. In this context, for example, we would like to emphasize for whom is interested in advanced topics the following aspects. Chapter 1 lays out a variety of notations, techniques and foundations basic to the material that is treated in this book. Chapter 2 introduces the powerful analytical instruments that, e.g., consist of principles of Tauberian theorems, including new results, with applications to convolution problems, evaluations of sequential procedures, renewal functions, and risk-efficient estimations. In this chapter we also consider problems of reconstructing the general distribution based on the distribution of some observed statistics. Chapter 3 shows certain nontrivial conclusions regarding the parametric likelihood ratios. Chapter 4 is developed to demonstrate a strong theoretical instrument based on martingales and their statistical applications, which include the martingale principle for testing statistical hypotheses and comparisons between the cumulative sum technique and the Shirayev–Roberts approach employed in change point detection policies. A part of material shown in Chapter 4 can be found only in this book. Chapter 5 can assist the statistician in developing and analyzing various Bayesian procedures. Chapter 8 provides the fundamental components for constructing unconventional statistical decision-making procedures with power one. Chapter 9 proposes novel approaches to examine, compare, and visualize properties of various statistical tests using correct p-value-based mechanisms. Chapter 10 introduces the empirical likelihood methodology. The theoretical propositions shown in Chapter 10 can lead to a quite mechanical and simple way to investigate properties of nonparametric likelihood-based statistical schemes. Several of these results can be found only in this book. In Chapter 14 one can discover interesting open problems. This book also provides software code based on both the R and SAS statistical software packages to exemplify the statistical methodological topics and their applied aspects.

Indeed, we focused essentially on developing a very informative textbook that introduces classical and novel statistical methods with respect to various biostatistical applications. Towards this end we employ our experience and relevant material obtained via our research and teaching activity across

10–15 years of biostatistical practice and training Master and PhD level students in the department of biostatistics. This book is intended for graduate students majoring in statistics, biostatistics, epidemiology, health-related sciences, and/or in a field where a statistics concentration is desirable, particularly for those who are interested in formal statistical mechanisms and their evaluations. In this context Chapters 1–10 and 12–14 provides teaching sources for a high level statistical theory course that can be taught in a statistical/biostatistical department. The presented material evolved in conjunction with teaching such a one-semester course at The New York State University at Buffalo. This course, entitled “Theory of Statistical Inference,” has belonged to a set of the four core courses required for our biostatistics PhD program.

This textbook delivers a “ready-to-go” well-structured product to be employed in developing advanced biostatistical courses. We offer lectures, homework questions, and their solutions, examples of midterm, final, and Ph.D. qualifying exams as well as examples of students’ projects.

One of the ideas regarding this book’s development is that we combine presentations of traditional applied and theoretical statistical methods with computationally extensive bootstrap type procedures that are relatively novel data-driven statistical tools. Chapter 11 is proposed to help instructors acquire a statistical course that introduces the Jackknife and Bootstrap methods. The focus is on the statistical functional as the key component of the theoretical developments with applied examples provided to illustrate the corresponding theory.

We strongly suggest to begin lectures by asking students to smile!

It is recommended to start each lecture class by answering students’ inquiries regarding the previously assigned homework problems. In this course, we assume that students are encouraged to present their work in class regarding individually tailored research projects (e.g., Chapter 14). In this manner, the material of the course can be significantly extended.

Our intent is not that this book competes with classical fundamental guides such as, e.g., Bickel and Doksum (2007), Borovkov (1998), and Serfling (2002). In our course, we encourage scholars to read the essential works of the world-renown authors. We aim to present different theoretical approaches that are commonly used in modern statistics and biostatistics to (1) analyze properties of statistical mechanisms; (2) compare statistical procedures; and (3) develop efficient (optimal) statistical schemes. Our target is to provide scholars research seeds to spark new ideas. Towards this end, we demonstrate open problems, basic ingredients in learning complex statistical notations and tools as well as advanced nonconventional methods, even in simple cases of statistical operations, providing “**Warning**” remarks to show potential difficulties related to the issues that were discussed.

Finally, we would like to note that this book attempts to represent a part of our life that definitely consists of mistakes, stereotypes, puzzles, and so on, that we all love. Thus our book cannot be perfect. We truly thank the reader

for his/her participation in our life! We hope that the presented material can play a role as prior information for various research outputs.

Albert Vexler
Alan D. Hutson



Authors

Albert Vexler, PhD, obtained his PhD degree in Statistics and Probability Theory from the Hebrew University of Jerusalem in 2003. His PhD advisor was Moshe Pollak, a fellow of the *American Statistical Association* and Marcy Bogen Professor of Statistics at Hebrew University. Dr. Vexler was a postdoctoral research fellow in the Biometry and Mathematical Statistics Branch at the National Institute of Child Health and Human Development (National Institutes of Health). Currently, Dr. Vexler is a tenured Full Professor at the State University of New York at Buffalo, Department of Biostatistics. Dr. Vexler has authored and co-authored various publications that contribute to both the theoretical and applied aspects of statistics in medical research. Many of his papers and statistical software developments have appeared in statistical/biostatistical journals, which have top-rated impact factors and are historically recognized as the leading scientific journals, and include: *Biometrics*, *Biometrika*, *Journal of Statistical Software*, *The American Statistician*, *The Annals of Applied Statistics*, *Statistical Methods in Medical Research*, *Biostatistics*, *Journal of Computational Biology*, *Statistics in Medicine*, *Statistics and Computing*, *Computational Statistics and Data Analysis*, *Scandinavian Journal of Statistics*, *Biometrical Journal*, *Statistics in Biopharmaceutical Research*, *Stochastic Processes and their Applications*, *Journal of Statistical Planning and Inference*, *Annals of the Institute of Statistical Mathematics*, *The Canadian Journal of Statistics*, *Metrika*, *Statistics*, *Journal of Applied Statistics*, *Journal of Nonparametric Statistics*, *Communications in Statistics*, *Sequential Analysis*, *The STATA Journal*; *American Journal of Epidemiology*, *Epidemiology*, *Paediatric and Perinatal Epidemiology*, *Academic Radiology*, *The Journal of Clinical Endocrinology & Metabolism*, *Journal of Addiction Medicine*, and *Reproductive Toxicology and Human Reproduction*.

Dr. Vexler was awarded National Institutes of Health (NIH) grants to develop novel nonparametric data analysis and statistical methodology. His research interests are related to the following subjects: receiver operating characteristic curves analysis; measurement error; optimal designs; regression models; censored data; change point problems; sequential analysis; statistical epidemiology; Bayesian decision-making mechanisms; asymptotic methods of statistics; forecasting; sampling; optimal testing; nonparametric tests; empirical likelihoods, renewal theory; Tauberian theorems; time series; categorical analysis; multivariate analysis; multivariate testing of complex hypotheses; factor and principal component analysis; statistical biomarker evaluations; and best combinations of biomarkers. Dr. Vexler is Associate Editor for *Biometrics* and *Journal of Applied Statistics*. These journals belong to the first cohort of academic literature related to the methodology of biostatistical and epidemiological research and clinical trials.

Alan D. Hutson, PhD, received his BA (1988) and MA (1990) in Statistics from the State University of New York (SUNY) at Buffalo. He then worked for Otsuka America Pharmaceuticals for two years as a biostatistician. Dr. Hutson then received his MA (1993) and PhD (1996) in Statistics from the University of Rochester. His PhD advisor was Professor Govind Mudholkar, a world-renown researcher in Statistics and Biostatistics. He was hired as a biostatistician at the University of Florida in 1996 as a Research Assistant Professor and worked his way to a tenured Associate Professor. He had several roles at the University of Florida including Interim Director of the Division of Biostatistics and Director of the General Clinical Research Informatics Core. Dr. Hutson moved to the University at Buffalo in 2002 as an Associate Professor and Chief of the Division of Biostatistics. He was the founding chair of the new Department of Biostatistics in 2003 and became a full professor in 2007. His accomplishments as Chair included the implementation of several new undergraduate and graduate degree programs and a substantial growth in the size and quality of the department faculty and students. In 2005, Dr. Hutson also became Chair of Biostatistics (now Biostatistics and Bioinformatics) at Roswell Park Cancer Institute (RPCI), was appointed Professor of Oncology, and became the Director of the Core Cancer Center Biostatistics Core. Dr. Hutson helped implement the new Bioinformatics Core at RPCI. He recently became the Deputy Group Statistician for the NCI national NRG cancer cooperative group. Dr. Hutson is Fellow of the *American Statistical Association*. He is Associate Editor of *Communications in Statistics*, Associate Editor of the *Sri Lankan Journal of Applied Statistics*, and is a New York State NYSTAR Distinguished Professor. He has membership on several data safety and monitoring boards and has served on several high level scientific review panels. He has over 200 peer-reviewed publications. In 2013, Dr. Hutson was inducted into the Delta Omega Public Health Honor Society, Gamma Lambda Chapter. His methodological work focuses on non-parametric methods for biostatistical applications as it pertains to statistical functionals. He has several years of experience in the design and analysis of clinical trials.

Subject Index

A

Absolute value, 2, 15, 16
 Adaptive CUSUM test statistic, 108, 109
 Adaptive Gaussian quadrature method, 147
 Adaptive maximum likelihood estimation, 105–109
 Adaptive maximum likelihood ratio, 107, 108
 Adaptive sequential designs, 181
 Almost sure convergence, 9, 10
 Anti-conservative approach, 135
A posteriori probabilities, 136
 Approximate confidence intervals, 264–265
 Approximate likelihood technique, 239
A priori information, 129
A priori probabilities, 136
 Area Under the ROC curve (AUC), 189–190, 202, 227
 estimated, 197
 improvement of, 366
 nonparametric approach, 192–193
 overestimation of, 196–197
 parametric approach, 190–192
 partial, 203, 226
 of student's *t*-test, 228
 Artificial likelihood-based techniques, 239
 ASN, *see* Average sample number
 Asymmetric distributions, 281
 Asymptotic approximations, 265
 Bayes factor, 142, 154
 benefits of, 146
 Laplace's method, 142
 Schwarz criterion, 145–147
 simulation and, 148–149
 variants on Laplace's method, 142–145
 Asymptotic properties of stopping time, 174–177
 At most one change-point model, 101
 AUC, *see* Area Under the ROC curve

Autoregressive errors, measurement model with, 80, 307
 Average run length (ARL) to false alarm, 112, 113
 Average sample number (ASN), 173
 Wald approximation, 173–174, 176

B

Bartlett correction, 245
 Bayes factor, 79, 99, 103, 124, 136, 145
 applications, 158–161
 computation, 137–142
 asymptotic approximations, *see* Asymptotic approximations
 importance sampling, 147
 samples from posterior distributions, 147–148
 simple Monte Carlo, 147
 decision-making rules, 153–158
 implications, 161–162
 principles, 132
 prior distributions, 149–153
 R code, 162–163
 Ville and Wald inequality, 209
 Bayesian approach, 129
 confidence interval estimation, 219, 220
 and empirical likelihood, 254–255
 equal-tailed confidence interval estimation, 221
 frequentist method *vs.*, 130
 highest posterior density confidence interval estimation, 221
 for hypothesis testing, 136, 151
 a priori information, 129
 Bayesian inference, 150
 best combination of multiple biomarkers, 363–364
 empirical, 364
 nonparametric, 255
 Bayesian information criterion (BIC), 145–146

- Bayesian methodology, 129
 advantages of, 131
 criticism of, 153
 information source, 133
 Bayesian priors, 31
 Bayes theorem, 59, 129, 136, 219, 253
 Berry–Esse'en inequality, 286
 Best linear combination (BLC) of
 biomarkers, 201–202
 Bias estimation, jackknife method,
 261–263
 BIC, *see* Bayesian information criterion
 “Big Oh” notation, 2, 10
 Biomarker(s), 54
 high density lipoprotein cholesterol,
 232
 in medical decision making, 231–232
 multiple, best combination of,
 201–202
 Bayesian inference, 363–364
 empirical Bayesian inference,
 364
 log-normal distribution, 364
 Biostatistical theory, 83
 Bivariate normal distribution, 60
 Bootstrap consistency, 285, 286
 Bootstrap method, 259–260, 266–268
 bias-corrected and accelerated
 interval, 297–299
 bootstrap-*t* percentile interval, 280,
 288–292
 confidence intervals, 279–284,
 291–303
 Edgeworth expansion technique, 279,
 284–288
 nonparametric simulation, 267–269
 resampling algorithms with SAS and
 R codes, 270–279
 Bootstrap replications, 260, 268–271, 273,
 275, 276
 Bootstrap tilting, 303–304
 Bootstrap-*t* percentile interval, 280,
 288–292
- C**
- Case-control studies, 164, 189, 231
 Cauchy distribution, 35, 57, 68
 Cauchy–Schwarz inequality, 11
 Central limit theorem (CLT), 174, 205,
 215, 310
 characteristic functions, 50–54
 Gaussian, 148
 maximum likelihood estimation,
 64–66
 for stopping time, 177–179
 Change of measure, 75
 Change point detection method
 retrospective, 100–105
 sequential, 109–113
 transformation of, *see* Shirayayev–
 Roberts technique
 Change point problems, 100, 101, 103,
 109, 363
 Characteristic exponent, 57
 Characteristic function
 algorithm, 40
 central limit theorem, 50–54
 distribution functions
 analytical forms of, 49–50
 reconstruction issues, 54–56
 elementary properties of, 32–34
 law of large numbers, 48–49
 one-to-one mapping, 34–39
 parametric approach, 56–57
 risk-efficient estimation, 46–47
 Tauberian theorems, 40–46
 Chebyshev's inequality, 11, 12, 48, 63, 68,
 92, 249
 Chi-square distribution, 181, 241, 365
 Classical retrospective hypothesis
 testing, 166
 Clinical pulmonary infection score
 (CPIS), 254
 Coefficient of variation, 268, 269
 Complex conjugate, 15, 32
 Complex variables, 14–16
 Composite estimation, of mean, 366–367
 Composite hypothesis, 130
 Conditional expectation, 85
 Conditional power computation, 182
 Conditional probability theory, 81
 Confidence band, 264
 Confidence interval estimation
 Bayesian approach, 219, 220
 equal-tailed, 221
 highest posterior density, 221, 222
 Confidence interval estimators, 66, 222

- Confidence intervals, 215
 - approximate, 264–265
 - bootstrap, 279–284, 291–303
 - constructions, 214, 216
 - definition, 263–264
 - estimation by SAS program, 289–291, 293–302
 - hypothesis testing, 222
 - posterior probabilities, 220
- Confidence sequences, 213, 216
 - advantage of, 214
 - for median, 215–217
- Conjugate distributions, 149
- Conjugate priors, 137, 149
- Consistency, 60
- Continuous mapping theorem, 9
- Continuous random variables, 4, 41, 108, 109
- Control limits, 110, 111
- Control statements, SAS, 26–28
- Convergence
 - almost sure, 9
 - associative properties, 10
 - in distribution, 9
 - in probability, 9
 - rate, 54
 - in r^{th} mean, 10
- Convolution, 33, 81, 86,
 - see also* Deconvolution
- Correct model-based likelihood formation, 80–82
- Course projects, examples of, 363–367
- Coverage probability, 66
- CPIS, *see* Clinical pulmonary infection score
- Credible set, 219
- Cumulative distribution functions, 152
- Cumulative sum (CUSUM) technique
 - change point detection, 101–102
 - modification of, 126–128
 - Shiryayev–Roberts *vs.*, 120–123
- D**
- Decision(-making) rule, 98, 133, 134, 256
 - Bayes factor, 153–158
 - likelihood ratio-based test, 69–71, 78, 130
 - most powerful, 135
 - null hypothesis rejection, 133
 - sequential probability ratio test, 170
- Deconvolution, 55
- Delta method, 190, 191
- Density function, 4–6
- Dirichlet integral, 37
- Discrete random variables, 4
- Distribution function, 3, 31, 43
 - analytical forms of, 49–50
 - characteristic function of, 32
 - continuous, 4
 - convergence in, 9, 10
 - empirical, 7, 188, 261, 266, 268
 - extensions and estimations, 56–57
 - Laplace transformations of, 31
 - of measurement errors, 55
 - one-to-one mapping, 34–39
 - parametric, 5–6
 - reconstruction issues, 54–56
 - of standard normal random variable, 66
 - sums of random variables, 32, 33
- DO-END statement, 27
- Doob's decomposition, 114, 116
- Doob's inequality, 92–93, 100, 205
- E**
- Econometrics, 83
- Edgeworth expansion technique, 279, 284–288
- Efron's formula, 195
- Elementary renewal theorem, 46
- Elicited prior, 133, 149
- EL methods, *see* Empirical likelihood methods
- Empirical Bayes(ian), 129
 - concept, 130, 141
 - inference, 364
 - nonparametric version, 255
- Empirical distribution function, 7, 188, 261, 266, 268
- Empirical likelihood functions, 222
- Empirical likelihood (EL) methods, 303
 - Bayesian approach and, 254–255
 - classical, 239–241
 - and parametric likelihood methods, 242, 253–254

- practical theoretical tools for
 - analyzing, 249–253
 - software packages, 241–242
 - as statistical analysis tool, 255–256
 - techniques for analyzing, 242–248
- Empirical likelihood ratio tests, 365–366
- Empirical likelihood ratio test statistic,
 - 241, 244–246, 256–257
- EPV, *see* Expected p -value
- Equal-tailed confidence interval
 - estimation, 66, 221
- Errors, *see also* Type I error; Type II error
 - autoregressive, 80
 - measurement, 55
 - Monte Carlo, 268
- Estimation, *see also* Overestimation
 - bias, 261–263
 - confidence interval, *see* Confidence interval estimation
 - James–Stein, 255, 364
 - maximum likelihood, *see* Maximum likelihood estimation
 - mean
 - composite, 366–367
 - trimmed, *see* Trimmed mean estimation
 - nonanticipating, 105–109, 124–125
 - risk-efficient, 46–47
 - unbiased, 167, 183, 273
- Euler’s formula, 15–16, 32, 36
- Exam questions, examples
 - final exams, 326–340
 - midterm exams, 326–340
 - qualifying exams, 326–340
- Expectation, 70
 - conditional, 85
 - definition of, 6
 - and integration, 6–8
 - law of, 86
- Expected bias
 - asymptotic approximation to, 195
 - of ROC curve, 196–197
- Expected p -value (EPV), 225,
 - see also* Partial expected p -value
- power and, 229–231
- ROC curve analysis, 226–227, 236–237
- of t -test, 233–234
- Expected significance level, 225
- Exponential distribution, 47, 132
 - density function for, 5, 6
- F**
 - False alarm, average run length to, 112, 113
 - False positive rate (FPR), 194, 195, 203
 - Fatou–Lebesgue theorem, 205
 - Fisher information, 64, 139–141
 - Fisher’s combination method, 181
 - Fisher’s z -transformation, 292
 - Flat prior, 135
 - FPR, *see* False positive rate
 - Frequentist approach, 129, 130, 133
 - Futility analysis, 168, 182
- G**
 - Gambling theory, 83
 - Gamma distribution, 140
 - analytical forms of, 49
 - density function, 5, 6, 138
 - Gaussian quadrature, 142, 147
 - Gaussian-type integral, 145
 - Geometric progression, sum formula
 - for, 42, 43
 - Goodness-of-fit tests, 307
 - Grouped jackknife approach, 262
 - Group sequential tests, 166, 179–181
- H**
 - Hessian matrix, 139–140, 143
 - Highest posterior density confidence interval estimation, 221, 222
 - Highly peaked integrand, 137
 - Hinchin’s form, law of large numbers, 48–49
 - Homework questions, examples of, 305–313
 - Hypothesis
 - composite, 130
 - formulation of, 69
 - nested, 143
 - null, *see* Null hypothesis
 - Hypothesis testing, 75
 - Bayesian approach for, 136, 151
 - classical, 166

composite statement, 96
 martingale principle, 93–99
 most powerful test, 98
 problems, 130

I

IF-THEN statement, 26
 Importance sampling, 147, 160
 Improper distribution, 150
 In control process, 110, 113, 114, 123
 Independent and identically distributed
 (iid) random variables
 distribution function of, 54, 56
 probability distribution, 4
 sum of, 65, 66, 173, 178
 in Ville and Wald inequality,
 209–213
 Indicator function, 7, 10–12, 70, 135, 150
 Inference
 AUC-based, 366
 Bayesian, 150, 255, 363–364
 empirical Bayesian, 364
 frequentist context, 129
 on incomplete bivariate data, 253
 ROC curve, 186–189
 INFILE statement, 24–25
 Infimum, 7
 Integrated most powerful test, 132–135
 Integration, expectation and, 6–8
 Invariant data transformations, 108
 Invariant statistics, 106
 Invariant transformations, 106
 Inversion theorem, 34–39

J

Jackknife method, 259
 bias estimation, 261–263
 confidence interval
 approximate, 264–265
 definition, 263–264
 homework questions, 312–313
 nonparametric estimate of variance,
 268
 variance estimation, 263
 variance stabilization, 265
 James–Stein estimation, 255, 364
 Jensen’s inequality, 14, 61, 87, 95

K

Kernel function, 192
 Khinchin’s form, law of large numbers,
 48–49
 Kullback–Leibler distance, 304

L

Lagrange method, 221, 240, 242
 Lagrange multiplier, 221, 240, 243,
 250, 256
 Laplace approximation, 142
 Laplace distribution, 35
 Laplace’s method, 142
 modification of, 148
 simulated version of, 149
 variants on, 142–145
 Laplace transformation
 of distribution functions, 31
 real-valued, 44
 Law of iterated expectation, 86
 Law of large numbers, 32, 48, 205
 Khinchin’s (Hinchin’s) form of, 48–49
 outcomes of, 51, 54
 strong, 10
 weak, 9
 Law of the iterated logarithm, 205,
 206–207, 214, 217
 Law of total expectation, 86
 LCL, *see* Lower control limit
 Least squares method, 61
 Lebesgue principle, 7
 Lebesgue–Stieltjes integral, 7, 8
 Likelihood function, 59, 60
 Likelihood method, 59–60
 artificial/approximate, 239
 correct model-based, 80
 intuition about, 61–63
 parametric, 78
 Likelihood ratio (LR), 69–73, 97–99, 170,
 see also Log-likelihood ratio;
 Maximum likelihood ratio
 Likelihood ratio test
 bias, 74
 decision rule, 69–71
 optimality of, 69, 71–72
 power, 70, 74
 type I error control, 100

- Likelihood ratio test statistic, 69, 73–75, 96
- Limits, 1–2, 8
- Linear regressions, empirical likelihood ratio tests for, 365–366
- “Little oh” notation, 2, 10
- Log-empirical likelihood ratios, 246–248
- Logistic regression, 185
 - change point problems in, 363
 - ROC curve analysis, 193–195
 - example, 198–201
 - expected bias of ROC, 196–197
 - retrospective and prospective ROC, 195–196
- Log likelihood function, 63, 79, 137–141
- Log-likelihood ratio, 61, 143, 144, 173, 310
- Log maximum likelihood, 64
- Log-normal distribution, 5–6
- Log-normally distributed values
 - composite estimation of mean, 366–367
 - of multiple biomarkers, 364
- Lower confidence band, 264
- Lower control limit (LCL), 111
- LR, *see* Likelihood ratio

- M**
- Magnitude of complex number, 15
- Mallow’s distance, 286–287
- Mann–Whitney statistic, 192, 198
- Markov chain Monte Carlo (MCMC) method, 148, 221
- Martingale-based arguments, 79
- Martingale transforms, 116–118, 121
- Martingale type statistics, 83
 - adaptive estimators, 105–109
 - change point detection, *see* Change point detection method
 - conditions for, 86
 - Doob’s inequality, 92–93
 - hypotheses testing, 93–99
 - likelihood ratios, 97–99
 - maximum likelihood ratio, 96–97
 - optional stopping theorem, 91
 - reverse martingale, 102, 108
 - σ -algebra, 85–86
 - stopping time, 88–90
 - terminologies, 84–90
 - type I error rate control, 100, 106–107
 - Wald’s lemma, 92
- Maximum likelihood estimation, 63–68, 131, 142, 183
- Maximum likelihood estimator (MLE), 47, 137, 143–144, 187, 191, 364, 366–367
 - adaptive, 105–109
 - example, 67–68
 - log-likelihood function, 137–139
 - as point estimator, 66
- Maximum likelihood ratio (MLR), 76–80, 96–97
- Maximum likelihood ratio test statistic, 76, 79, 96, 105
- MCMC method, *see* Markov chain Monte Carlo method
- Mean
 - composite estimation of, 366–367
 - r^{th} mean, convergence in, 10
- Measurability, 85
- Measurable random variable, 86
- Measurement errors, 32, 54–56, 363
- Measurement model, with
 - autoregressive errors, 80, 307
- Metropolis–Hastings algorithm, 148
- Minus maximum log-likelihood ratio test statistic, 246–248
- MLE, *see* Maximum likelihood estimator
- Model selection, 146
- Modulus of complex number, 15
- Monte Carlo (MC)
 - approximations, 102, 106, 229, 232
 - coverage probabilities, 223
 - error, 268
 - Markov chain, 148
 - resampling, 304
 - roulette game, 90
 - simplest, 147
- Monte Carlo power, 54, 198–199, 309
- Monte Carlo simulation, 113, 177, 267–268, 308, 310
 - principles of, 51–54
 - Shiryayev–Roberts technique, 124–128
- Most powerful decision rule, 135
- Most powerful test
 - of hypothesis testing, 98

integrated, 132–135
 likelihood ratio test, 69, 71
 Most probable position, 59
 Multiple biomarkers, best combination
 of, 201–202
 Bayesian inference, 363–364
 empirical Bayesian inference, 364
 log-normal distribution, 364
 Multitroot procedure, 191
 “mvnorm” command, 21–22

N

Nested hypothesis, 143
 Newton–Raphson method, 53, 68
 Neyman–Pearson lemma, 69, 78, 225, 231
 Neyman–Pearson testing concept, 69,
 73, 96, 98
 Nonanticipating estimation
 maximum likelihood estimation,
 105–109
 Monte Carlo simulation, 124–125
 Non-asymptotic upper bound of
 renewal function, 42
 Noninformative prior, 149–150
 Nonparametric approach
 area under the ROC curve, 192–193,
 198
 best linear combination of
 biomarkers, 202
 Nonparametric estimation of ROC
 curve, 187–188
 Nonparametric method
 Monte Carlo simulation scheme, 128
 Shiryayev–Roberts technique,
 123–124
 Nonparametric simulation, 267–269
 Nonparametric testing, 308
 Nonparametric tilting, *see* Bootstrap
 tilting
 Not a number (NaN) values, 18
 Null hypothesis, 97
 Bayes factor, 153
 CUSUM test statistic, 102
 empirical likelihood function, 246
 p -value, 161–162, 224, 225
 rejection of, 73, 79, 102, 130, 158, 170,
 222
 Numerical integration, 137, 142, 146

O

One-to-one mapping, 31, 34–39
 Online change point problem, 109
 Optimality, 69, 221
 likelihood argument of, 75
 of statistical tests, 71–72
 “Optim” procedure, 191
 Optional stopping theorem, 91
 Out of control process, 110, 111, 114, 123
 Overestimation, 194–195
 area under the ROC curve, 196–197

P

p -value, 76, 97, 132, 223–226, 237
 expected, *see* Expected p -value
 formulation of, 161–162
 null hypothesis, 161–162, 224, 225
 prediction intervals for, 225
 stochastic aspect of, 225
 Parametric approach
 area under the ROC curve, 190–192
 best linear combination of
 biomarkers, 202
 Parametric characteristic function,
 56–57
 Parametric distribution functions, 5–6
 Parametric likelihood (PL) method, 78,
 255
 empirical likelihood and, 253–254
 Parametric sequential procedures, 167
 Parametric statistical procedure, 167
 Partial AUC (pAUC), 203, 226
 Partial expected p -value (pEPV), 226,
 230, 231, 234–235
 pAUC, *see* Partial AUC
 Penalized empirical likelihood
 estimation, 366
 pEPV, *see* Partial expected p -value
 Percentile confidence interval, 280–282,
 288–291
 Pivot, 264
 PL method, *see* Parametric likelihood
 method
 Point estimator, 66
 Pooling design, 55–56
 Posterior distribution, 131, 146–149, 155
 Posterior empirical Bayes (PEB), 364

Posterior odds, 136, 161
 Post-sequential analysis, 182–183
 Power, 53
 calculation, 54
 expected p -value and, 229–231
 of likelihood ratio test, 70, 74
 Prior (probability) distribution, 131,
 149–153, 159–160, 219, 255
 Probability, convergence in, 9, 10
 Probability distribution, 3–4, 149–153
 Probability mass function, 4
 Probability theory, 3, 83, 86
 Probit regression model, 159, 160
 Problematic function, 7
 PROC IML, 29, 275, 277–279
 PROC MEANS
 procedure, 28–29
 WEIGHT statement, 273
 PROC SURVEYSELECT, 270, 274, 277
 PROC UNIVARIATE, 281–282, 289
 Product type chain rule form, 62

Q

Quasi-likelihood method, 239

R

Radon–Nikodym derivative, 75
 Radon–Nikodym theorem, 75
 RANBIN function, 270, 271
 Random variables, 2–3
 binomial, 270–271
 continuous, 4, 41, 108, 109
 discrete, 4
 measurable, 86
 Rasch model, 79
 Rate of convergence, 54
 R code, 17, 53–54, 57, 111, 112, 138–139,
 151, 157, 162–163, 170–172,
 176–177, 186, 189, 199–200, 229,
 232–233, 248, 270–279, 284, 303
 Receiver operating characteristic (ROC)
 curve analysis, 185, 364
 area under the ROC curve, 189–193
 in cardiovascular research, 185, 186
 case-control studies, 189
 curve inference, 186–189

 expected bias of, 196–197
 expected p -value and, 225, 226–227
 laboratory diagnostic tests, 185
 logistic regression and, 193–201
 nonparametric estimation of, 187–188
 student's t -test, 227, 228
 Regularity conditions, 63, 64, 66, 68, 76,
 77, 264
 Remainder term, 11, 12, 14, 64, 65, 77,
 145, 249, 287
 Renewal function, 40–46, 92
 Renewal theory, 31, 41, 178
 Replications (Bootstrap), 260, 268–271,
 273, 275, 276
 Representative method, 99, 132
 Resampling algorithms, 270–279
 Restricted maximum likelihood
 estimate (REML), 160
 Retrospective change point detection,
 100–101
 cumulative sum technique, 101–102
 Shiryayev–Roberts technique, 102–105
 Retrospective error rate, 195
 Retrospective mechanism, 88
 Retrospective studies, 165
 Reverse martingale, 102, 108
 Riemann–Stieltjes integral, 6, 8
 Risk-efficient estimation, 46–47
 Risk function, 46, 47, 71
 R software, 16, *see also* SAS software
 comparison operators, 20
 data manipulation, 18–21
 input window, 16–17
 printing data in, 21–22
 r^{th} mean, convergence in, 10

S

Sample size
 calculation, 54, 152
 expected, 47, 168, 169, 173, 218
 Sampling
 importance, 147, 160
 techniques, 149
 SAS/IML, 278
 SAS software, 16, 22
 comments in, 23–24
 comparison operators, 27

- control statements, 26–28
- data import, 24–25
- data manipulation, 25–28
- data printing, 28
- interface, 22, 23
- macros, 275–276
- numeric functions, 26
- resampling algorithms with, 270–279
- summarizing data in, 28–29
- trimmed mean and confidence
 - interval estimation, 289–291, 293–302
- variables and data sets, 23
- Schwarz criterion, 145–147, 154
- Semi-parametric confidence interval, 282
- Sequential analysis methodology
 - adaptive sequential designs, 181
 - advantages, 166
 - central limit theorem for stopping time, 177–179
 - futility analysis, 182
 - group sequential tests, 179–181
 - post-sequential analysis, 182–183
 - regulated trials, 166–167
 - sequential element, 165
 - sequential probability ratio test, 169–177
 - two-stage designs, 167–169
- Sequential change point detection, 109–113
- Sequential element, 165
- Sequential methods, 167, 179
- Sequential probability ratio test (SPRT), 169, 309–310
 - asymptotic properties of stopping time, 174–177
 - goal of, 170
 - sampling process, 171
 - stopping boundaries, 172–173
 - Wald approximation to average sample number, 173–174
- Sequential procedures, 83
 - guidelines for, 166
 - testing, 167
- Sequential trial, 166, 167, 183
- Shapiro–Wilk test for normality, 167, 183
- Shewhart chart (X-bar chart), 111
- Shiryayev–Roberts (SR) technique
 - change point detection, 102–105, 113
 - cumulative sum *vs.*, 120–123
 - method, 115–120
 - Monte Carlo simulation study, 124–128
 - motivation, 114–115
 - nonparametric example, 123–124
- σ -algebra, 85–86
- Simon’s two-stage designs, 168–169
- Simple representative method, 99, 132
- Simplest Monte Carlo integration, 147
- Skewed distributions, 5
- Slutsky’s theorem, 77
- Smooth statistics, 263, 280
- SPRT, *see* Sequential probability ratio test
- SR technique, *see* Shiryayev–Roberts technique
- Stabilization, variance, 265
- Standardizing decision-making policies, 223
- Statistical decision-making theory, 69
- Statistical functionals, 261, 266–268, 287, 303
- Statistical hypotheses testing,
 - martingale principle, 93–99
- Statistical literature, 59
- Statistical sequential schemes, 88, 167, 178
- Statistical significance, 223
- Stochastic curtailment approach, 182
- Stochastic processes analysis, 83
- Stopping times (rules), 41, 47, 83, 88–89, 92, 112, 124, 126, 173
 - asymptotic properties, 174–177
 - central limit theorem, 177–179
 - Ville and Wald inequality, 208
- Student’s *t*-test, 227–229
- Study designs, 54, 55
- Submartingale
 - condition, 86–87
 - H_0 -submartingale, 97, 114
 - H_1 -submartingale, 95, 96, 103
 - P_∞ -submartingale, 115, 116, 121, 122
- Sufficiency, notion of, 60
- Sum of iid random variables, 65, 66, 173, 178

Supermartingale, 86–87
 Supremum, 7, 106
 Symmetric distribution, 179, 211, 281, 287

T

Tauberian theorems, 32, 40–46
 Tauberian-type propositions, 44–46
 Taylor's theorem, 12–14, 43, 45, 64, 65, 191
 TBARS measurement, 311–313
 Test statistic
 adaptive cumulative sum, 108, 109
 adaptive maximum likelihood ratio, 107, 108
 definition of, 99
 empirical likelihood ratio, 241, 244–246, 256–257
 likelihood ratio, 69, 73–75, 96
 maximum likelihood ratio, 76, 79, 96, 105
 minus maximum log-likelihood ratio, 246
 Test with power one, 217–218
 “Tilde/twiddle” notation, 2
 Training samples, 194
 Trimmed mean estimation
 PROC UNIVARIATE, 289–290
 SAS program, 289–291, 293–302
 Trimmed mean statistic, 280–282
 Trivial inequality, 70
 True positive rate, 194, 195
t-tests, Wilcoxon rank-sum test *vs.*, 231–236
 Two-sample nonparametric testing, 308–309
 Two schools of thought, 129
 Two-stage acceptance sampling plan, 168
 Two-stage designs, sequential analysis, 167–169
 Type I error rate(s), 53, 167
 empirical likelihood ratio test, 242
 of maximum likelihood ratio test, 76
 notion of, 129
 probability of, 70, 73

Simon's two-stage designs, 168
 in statistical decision-making, 69, 72
 test with power one, 217–218
 Type I error rate control, 100, 106–107, 134, 145, 156, 207, 365–366
 Type II error probability, 73, 74
 test with power one, 218
 Type II error rate, 168
 Simon's two-stage designs, 168–169

U

Unbiased statistical test, 230
 Uniroot function, 52, 68, 175, 312
 Unit step function, 266
 Upper control limit (UCL), 111
 U-statistic theory, 192, 198

V

Validation samples, 194
 VAP, *see* Ventilator-associated pneumonia
 Variance, 155
 Bayes factor *vs.*, 160, 161
 estimation, 263, 268
 population, 262
 stabilization, 265
 Variants on Laplace's method, 142–145
 Ventilator-associated pneumonia (VAP), 254
 Ville and Wald inequality, 205–206, 208
 Bayes factor, 209
 distribution function, 210–211
 statistical significance, 213–218
 in sums of iid observations, 209–213
 symmetric distribution function, 211–213

W

Wald approximation to average sample number, 173–174, 176
 Wald martingale, 307
 Wald's lemma, 44, 92
 WEIGHT statement, 273
 Welch's *t*-test, 227–229

Wilcoxon rank-sum test, 225, 231–236
Wilks' theorem, 76, 78, 106, 241, 245,
365

X

X-chart, 110

Y

Youden index, 202, 203

Z

Z-interval, 265

