SPRINGER TEXTS IN STATISTICS

# Theory of Point Estimation

SECOND EDITION

# 点估计理论

第2版

E.L. Lehmann
George Casella

Springer

世界图书出版公司
www.wpcbj.com.cn

E.L. Lehmann George Casella

# Theory of Point Estimation

Second Edition

Springer

**图书在版编目（CIP）数据**

点估计理论：第2版 = Theory of Point Estimation 2nd ed.：英文/（美）瓦多莱曼（Lehmann，E. L.）著．—2版．—北京：世界图书出版公司北京公司，2009. 5
ISBN 978-7-5100-0484-1

Ⅰ．点… Ⅱ．瓦… Ⅲ．点估计—教材—英文 Ⅳ．O211. 67

中国版本图书馆CIP数据核字（2009）第055595号

| | |
|---|---|
| **书　　名：** | Theory of Point Estimation 2nd ed. |
| **作　　者：** | E. L. Lehmann & George Casella |
| **中 译 名：** | 点估计理论 第2版 |
| **责任编辑：** | 高蓉　刘慧 |
| **出 版 者：** | 世界图书出版公司北京公司 |
| **印 刷 者：** | 三河国英印务有限公司 |
| **发　　行：** | 世界图书出版公司北京公司（北京朝内大街137号　100010） |
| **联系电话：** | 010-64021602，010-64015659 |
| **电子信箱：** | kjb@ wpcbj. com. cn |
| **开　　本：** | 24开 |
| **印　　张：** | 26 |
| **版　　次：** | 2009年05月 |
| **版权登记：** | 图字：01-2009-2044 |
| **书　　号：** | 978-7-5100-0484-1/O·699　　**定　　价：** 70. 00元 |

# *Springer Texts in Statistics*

## Springer Texts in Statistics

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

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

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



Library of Congress Cataloging-in-Publication Data
Lehmann, E.L. (Erich Leo), 1917–
Theory of point estimation.—2nd ed./ E.L. Lehmann, George Casella
p. cm. — (Springer texts in statistics)
Includes bibliographical references and index.
ISBN 0-387-98502-6 (hardcover : alk. paper)
1. Fix-point estimation. I. Casella, George. II. Title.
III. Series.
QA276.8 .L43 1998
519.5′44—dc21 98-16687

ISBN-10: 0-387-98502-6 e-ISBN-10: 0-387-22728-8
ISBN-13: 978-0-387-98502-2 e-ISBN-13: 978-0-387-22728-3



12 11 10 9 8

To our children

Stephen, Barbara, and Fia
ELL

Benjamin and Sarah
GC

# Preface to the Second Edition

Since the publication in 1983 of *Theory of Point Estimation*, much new work has made it desirable to bring out a second edition. The inclusion of the new material has increased the length of the book from 500 to 600 pages; of the approximately 1000 references about 25% have appeared since 1983.

The greatest change has been the addition to the sparse treatment of Bayesian inference in the first edition. This includes the addition of new sections on Equivariant, Hierarchical, and Empirical Bayes, and on their comparisons. Other major additions deal with new developments concerning the information inequality and simultaneous and shrinkage estimation. The Notes at the end of each chapter now provide not only bibliographic and historical material but also introductions to recent development in point estimation and other related topics which, for space reasons, it was not possible to include in the main text. The problem sections also have been greatly expanded. On the other hand, to save space most of the discussion in the first edition on robust estimation (in particular L, M, and R estimators) has been deleted. This topic is the subject of two excellent books by Hampel et al (1986) and Staudte and Sheather (1990). Other than subject matter changes, there have been some minor modifications in the presentation. For example, all of the references are now collected together at the end of the text, examples are listed in a Table of Examples, and equations are references by section and number within a chapter and by chapter, section and number between chapters.

The level of presentation remains the same as that of TPE. Students with a thorough course in theoretical statistics (from texts such as Bickel and Doksum 1977 or Casella and Berger 1990) would be well prepared. The second edition of TPE is a companion volume to "Testing Statistical Hypotheses, Second Edition (TSH2)." Between them, they provide an account of classical statistics from a unified point of view.

Many people contributed to TPE2 with advice, suggestions, proofreading and problem-solving. We are grateful to the efforts of John Kimmel for overseeing this project; to Matt Briggs, Lynn Eberly, Rich Levine and Sam Wu for proofreading and problem solving, to Larry Brown, Anirban DasGupta, Persi Diaconis, Tom DiCiccio, Roger Farrell, Leslaw Gajek, Jim Hobert, Chuck McCulloch, Elias Moreno, Christian Robert, Andrew Rukhin, Bill Strawderman and

Larry Wasserman for discussions and advice on countless topics, and to June Meyermann for transcribing most of TPE to LaTeX. Lastly, we thank Andy Scherrer for repairing the near-fatal hard disk crash and Marty Wells for the almost infinite number of times he provided us with needed references.

E.L. Lehmann
Berkeley, California

George Casella
Ithaca, New York

March 1998

# Preface to the First Edition

This book is concerned with point estimation in Euclidean sample spaces. The first four chapters deal with exact (small-sample) theory, and their approach and organization parallel those of the companion volume, *Testing Statistical Hypotheses* (TSH). Optimal estimators are derived according to criteria such as unbiasedness, equivariance, and minimaxity, and the material is organized around these criteria. The principal applications are to exponential and group families, and the systematic discussion of the rich body of (relatively simple) statistical problems that fall under these headings constitutes a second major theme of the book.

A theory of much wider applicability is obtained by adopting a large sample approach. The last two chapters are therefore devoted to large-sample theory, with Chapter 5 providing a fairly elementary introduction to asymptotic concepts and tools. Chapter 6 establishes the asymptotic efficiency, in sufficiently regular cases, of maximum likelihood and related estimators, and of Bayes estimators, and presents a brief introduction to the local asymptotic optimality theory of Hajek and LeCam. Even in these two chapters, however, attention is restricted to Euclidean sample spaces, so that estimation in sequential analysis, stochastic processes, and function spaces, in particular, is not covered.

The text is supplemented by numerous problems. These and references to the literature are collected at the end of each chapter. The literature, particularly when applications are included, is so enormous and spread over the journals of so many countries and so many specialties that complete coverage did not seem feasible. The result is a somewhat inconsistent coverage which, in part, reflects my personal interests and experience.

It is assumed throughout that the reader has a good knowledge of calculus and linear algebra. Most of the book can be read without more advanced mathematics (including the sketch of measure theory which is presented in Section 1.2 for the sake of completeness) if the following conventions are accepted.

1. A central concept is that of an integral such as $\int f\, dP$ or $\int f\, d\mu$. This covers both the discrete and continuous case. In the discrete case $\int f\, dP$ becomes $\Sigma f(x_i)P(x_i)$ where $P(x_i) = P(X = x_i)$ and $\int f\, d\mu$ becomes $\Sigma f(x_i)$. In the continuous case, $\int f\, dP$ and $\int f\, d\mu$ become, respectively, $\int f(x)p(x)\, dx$ and $\int f(x)\, dx$. Little is

lost (except a unified notation and some generality) by always making these substitutions.

2. When specifying a probability distribution $P$, it is necessary to specify not only the sample space $\mathcal{X}$, but also the class $\mathcal{C}$ of sets over which $P$ is to be defined. In nearly all examples $\mathcal{X}$ will be a Euclidean space and $\mathcal{C}$ a large class of sets, the so-called Borel sets, which in particular includes all open and closed sets. The references to $\mathcal{C}$ can be ignored with practically no loss in the understanding of the statistical aspects.

A forerunner of this book appeared in 1950 in the form of mimeographed lecture notes taken by Colin Blyth during a course I taught at Berkeley; they subsequently provided a text for the course until the stencils gave out. Some sections were later updated by Michael Stuart and Fritz Scholz. Throughout the process of converting this material into a book, I greatly benefited from the support and advice of my wife, Juliet Shaffer. Parts of the manuscript were read by Rudy Beran, Peter Bickel, Colin Blyth, Larry Brown, Fritz Scholz, and Geoff Watson, all of whom suggested many improvements. Sections 6.7 and 6.8 are based on material provided by Peter Bickel and Chuck Stone, respectively. Very special thanks are due to Wei-Yin Loh, who carefully read the complete manuscript at its various stages and checked all the problems. His work led to the correction of innumerable errors and to many other improvements. Finally, I should like to thank Ruth Suzuki for her typing, which by now is legendary, and Sheila Gerber for her expert typing of many last-minute additions and corrections.

E.L. Lehmann
Berkeley, California,

March 1983

# Table of Notation

The following notation will be used throughout the book.
We present this list for easy reference.

| Quantity | Notation | Comment |
|---|---|---|
| Random variable | $X, Y,$ | uppercase |
| Sample space | $\mathcal{X}, \mathcal{Y}$ | uppercase script Roman letters |
| Parameter | $\theta, \lambda$ | lowercase Greek letters |
| Parameter space | $\Theta, \Omega$ | uppercase script Greek letters |
| Realized values (data) | $x, y$ | lowercase |
| Distribution function (cdf) | $F(\mathbf{x}), F(\mathbf{x}\vert\theta), P(\mathbf{x}\vert\theta)$ $F_\theta(x), P_\theta, (x)$ | continuous or discrete |
| Density function (pdf) | $f(\mathbf{x}), f(\mathbf{x}\vert\theta), p(\mathbf{x}\vert\theta)$ $f_\theta(x), P_\theta(x)$ | notation is "generic", i.e., don't assume $f(x\vert y) = f(x\vert z)$ |
| Prior distribution | $\Lambda(\gamma), \Lambda(\gamma\vert\lambda)$ | |
| Prior density | $\pi(\gamma), \pi(\gamma\vert\lambda)$ | may be improper |
| Probability triple | $(\mathcal{X}, \mathcal{P}, \mathcal{B})$ | sample space, probability distribution, and sigma-algebra of sets |

| Quantity | Notation | Comment |
|---|---|---|
| Vector | $\mathbf{h} = (h_1, \ldots, h_n) = \{h_i\}$ | boldface signifies vectors |
| Matrix | $H = \{h_{ij}\} = \|\|h_{ij}\|\|$ | uppercase signifies matrices |
| Special matrices and vectors | $I$<br>$\mathbf{1}$<br>$J = \mathbf{11}'$ | Identity matrix<br>vector of ones<br>matrix of ones |
| Dot notation | $h_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} h_{ij}$ | average across the dotted subscript |
| Gradient | $\nabla h(\mathbf{x}) = \left(\frac{\partial}{\partial x_1}h(\mathbf{x}), \ldots, \frac{\partial}{\partial x_n}h(\mathbf{x})\right) = \left\{\frac{\partial}{\partial x_i}h(\mathbf{x})\right\}$ | vector of partial derivatives |
| Hessian | $\nabla\nabla h(\mathbf{x}) = \left\{\frac{\partial^2}{\partial x_i \partial x_j}h(\mathbf{x})\right\}$ | matrix of partial second derivatives |
| Jacobian | $\left\{\frac{\partial}{\partial x_j}h_i(\mathbf{x})\right\}$ | matrix of derivatives |
| Laplacian | $\sum_i \frac{\partial^2}{\partial x_i^2}h(\mathbf{x})$ | sum of second derivatives |
| Euclidean norm | $\|\mathbf{x}\|$ | $(\Sigma x_i^2)^{1/2}$ |
| Indicator function | $I_A(x)$, $I(x \in A)$ or $I(x < a)$ | equals 1 if $x \in A$, 0 otherwise |
| Big "Oh," little "oh" | $O(n)$, $o(n)$ or $O_p(n)$, $o_p(n)$ | As $n \to \infty$ $\frac{O(n)}{n} \to$ constant, $\frac{o(n)}{n} \to 0$ subscript $p$ denotes *in probability* |

# Contents

# List of Tables