"Non-asymptotic, high-dimensional theory is critical for modern statistics and machine learning. This book is unique in providing a crystal clear, complete, and unified treatment of the area. With topics ranging from concentration of measure to graphical models, the author weaves together probability theory and its applications to statistics. Ideal for graduate students and researchers. This will surely be the standard reference on the topic for many years."

— Larry Wasserman, *Carnegie Mellon University*

"Martin Wainwright brings his large box of analytical power tools to bear on the problems of the day—the analysis of models for wide data. A broad knowledge of this new area combines with his powerful analytical skills to deliver this impressive and intimidating work—bound to be an essential reference for all the brave souls that try their hand."

— Trevor Hastie, *Stanford University*

"This book provides an excellent treatment of perhaps the fastest growing area within high-dimensional theoretical statistics—non-asymptotic theory that seeks to provide probabilistic bounds on estimators as a function of sample size and dimension. It offers the most thorough, clear, and engaging coverage of this area to date, and is thus poised to become the definitive reference and textbook on this topic."

— Genevera Allen, *Rice University*

"Statistical theory and practice have undergone a renaissance in the past two decades, with intensive study of high-dimensional data analysis. No researcher has deepened our understanding of high-dimensional statistics more than Martin Wainwright. This book brings the signature clarity and incisiveness of his published research into book form. It will be a fantastic resource for both beginning students and seasoned researchers, as the field continues to make exciting breakthroughs."

— John Lafferty, *Yale University*

"This is an outstanding book on high-dimensional statistics, written by a creative and celebrated researcher in the field. It gives comprehensive treatments of many important topics in statistical machine learning and, furthermore, is self-contained, from introductory material to the most up-to-date results on various research frontiers. This book is a must-read for those who wish to learn and to develop modern statistical machine theory, methods and algorithms."

— Jianqing Fan, *Princeton University*

"This book provides an in-depth mathematical treatment and methodological intuition for high-dimensional statistics. The main technical tools from probability theory are carefully developed and the construction and analysis of statistical methods and algorithms for high-dimensional problems are presented in an outstandingly clear way. Martin Wainwright has written a truly exceptional, inspiring, and beautiful masterpiece!"

— Peter Bühlmann, *ETH Zurich*

"This new book by Martin Wainwright covers modern topics in high-dimensional statistical inference, and focuses primarily on explicit non-asymptotic results related to sparsity and non-parametric estimation. This is a must-read for all graduate students in mathematical statistics and theoretical machine learning, both for the breadth of recent advances it covers and the depth of results which are presented. The exposition is outstandingly clear, starting from the first introductory chapters on the necessary probabilistic tools. Then, the book covers state-of-the-art advances in high-dimensional statistics, with always a clever choice of results which have the perfect mix of significance and mathematical depth."

— Francis Bach, *INRIA Paris*

"Wainwright's book on those parts of probability theory and mathematical statistics critical to understanding of the new phenomena encountered in high dimensions is marked by the clarity of its presentation and the depth to which it travels. In every chapter he starts with intuitive examples and simulations which are systematically developed either into powerful mathematical tools or complete answers to fundamental questions of inference. It is not easy, but elegant and rewarding whether read systematically or dipped into as a reference."

— Peter Bickel, *UC Berkeley*

High-Dimensional Statistics

Recent years have witnessed an explosion in the volume and variety of data collected in all scientific disciplines and industrial settings. Such massive data sets present a number of challenges to researchers in statistics and machine learning. This book provides a self-contained introduction to the area of high-dimensional statistics, aimed at the first-year graduate level. It includes chapters that are focused on core methodology and theory—including tail bounds, concentration inequalities, uniform laws and empirical process, and random matrices—as well as chapters devoted to in-depth exploration of particular model classes—including sparse linear models, matrix models with rank constraints, graphical models, and various types of non-parametric models.

With hundreds of worked examples and exercises, this text is intended both for courses and for self-study by graduate students and researchers in statistics, machine learning, and related fields who must understand, apply, and adapt modern statistical methods suited to large-scale data.

MARTIN J. WAINWRIGHT is a Chancellor's Professor at the University of California, Berkeley, with a joint appointment between the Department of Statistics and the Department of Electrical Engineering and Computer Sciences. His research lies at the nexus of statistics, machine learning, optimization, and information theory, and he has published widely in all of these disciplines. He has written two other books, one on graphical models together with Michael I. Jordan, and one on sparse learning together with Trevor Hastie and Robert Tibshirani. Among other awards, he has received the COPSS Presidents' Award, has been a Medallion Lecturer and Blackwell Lecturer for the Institute of Mathematical Statistics, and has received Best Paper Awards from the NIPS, ICML, and UAI conferences, as well as from the IEEE Information Theory Society.

CAMBRIDGE SERIES IN STATISTICAL AND PROBABILISTIC MATHEMATICS

This series of high-quality upper-division textbooks and expository monographs covers all aspects of stochastic applicable mathematics. The topics range from pure and applied statistics to probability theory, operations research, optimization, and mathematical programming. The books contain clear presentations of new developments in the field and also of the state of the art in classical methods. While emphasizing rigorous treatment of theoretical methods, the books also contain applications and discussions of new techniques made possible by advances in computational practice.

A complete list of books in the series can be found at www.cambridge.org/statistics.
Recent titles include the following:

23. *Applied Asymptotics,* by A. R. Brazzale, A. C. Davison and N. Reid
24. *Random Networks for Communication,* by Massimo Franceschetti and Ronald Meester
25. *Design of Comparative Experiments,* by R. A. Bailey
26. *Symmetry Studies,* by Marlos A. G. Viana
27. *Model Selection and Model Averaging,* by Gerda Claeskens and Nils Lid Hjort
28. *Bayesian Nonparametrics,* edited by Nils Lid Hjort *et al.*
29. *From Finite Sample to Asymptotic Methods in Statistics,* by Pranab K. Sen, Julio M. Singer and Antonio C. Pedrosa de Lima
30. *Brownian Motion,* by Peter Mörters and Yuval Peres
31. *Probability: Theory and Examples (Fourth Edition),* by Rick Durrett
33. *Stochastic Processes,* by Richard F. Bass
34. *Regression for Categorical Data,* by Gerhard Tutz
35. *Exercises in Probability (Second Edition),* by Loïc Chaumont and Marc Yor
36. *Statistical Principles for the Design of Experiments,* by R. Mead, S. G. Gilmour and A. Mead
37. *Quantum Stochastics,* by Mou-Hsiung Chang
38. *Nonparametric Estimation under Shape Constraints*, by Piet Groeneboom and Geurt Jongbloed
39. *Large Sample Covariance Matrices and High-Dimensional Data Analysis,* by Jianfeng Yao, Shurong Zheng and Zhidong Bai
40. *Mathematical Foundations of Infinite-Dimensional Statistical Models,* by Evarist Giné and Richard Nickl
41. *Confidence, Likelihood, Probability,* by Tore Schweder and Nils Lid Hjort
42. *Probability on Trees and Networks*, by Russell Lyons and Yuval Peres
43. *Random Graphs and Complex Networks (Volume 1)*, by Remco van der Hofstad
44. *Fundamentals of Nonparametric Bayesian Inference*, by Subhashis Ghosal and Aad van der Vaart
45. *Long-Range Dependence and Self-Similarity*, by Vladas Pipiras and Murad S. Taqqu
46. *Predictive Statistics*, by Bertrand S. Clarke and Jennifer L. Clarke
47. *High-Dimensional Probability,* by Roman Vershynin
48. *High-Dimensional Statistics,* by Martin J. Wainwright
49. *Probability: Theory and Examples (Fifth Edition),* by Rick Durrett

High-Dimensional Statistics

A Non-Asymptotic Viewpoint

Martin J. Wainwright
University of California, Berkeley

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108498029
DOI: 10.1017/9781108627771

First published 2019

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Wainwright, Martin (Martin J.), author.
Title: High-dimensional statistics : a non-asymptotic viewpoint / Martin J.
Wainwright (University of California, Berkeley).
Description: Cambridge ; New York, NY : Cambridge University Press, 2019. |
Series: Cambridge series in statistical and probabilistic mathematics ; 48 |
Includes bibliographical references and indexes.
Identifiers: LCCN 2018043475 | ISBN 9781108498029 (hardback)
Subjects: LCSH: Mathematical statistics–Textbooks. | Big data.
Classification: LCC QA276.18 .W35 2019 | DDC 519.5–dc23
LC record available at https://lccn.loc.gov/2018043475

ISBN 978-1-108-49802-9 Hardback

Cambridge University Press
978-1-108-49802-9 — High-Dimensional Statistics
Martin J. Wainwright
Frontmatter
More Information

List of chapters

Contents

Illustrations

Acknowledgements

This book would not exist without the help of many people in my life. I thank my parents, John and Patricia, for their nurture and support over the years, and my family, Haruko, Hana, Mina, and Kento for bringing daily joy. In developing my own understanding of and perspective on the broad area of high-dimensional statistics, I have been fortunate to interact with and learn from many wonderful colleagues, both here at Berkeley and elsewhere. For valuable discussions, insights and feedback, I would like to thank, among others, Bryon Aragam, Alex d'Asprémont, Francis Bach, Peter Bickel, Peter Bühlmann, Tony Cai, Emmanuel Candès, Constantine Caramanis, David Donoho, Noureddine El Karoui, Jianqing Fan, Aditya Guntuboyina, Trevor Hastie, Iain Johnstone, Michael Jordan, John Lafferty, Eliza Levina, Zongming Ma, Nicolai Meinshausen, Andrea Montanari, Axel Munk, Richard Nickl, Eric Price, Philippe Rigollet, Alessandro Rinaldo, Richard Samworth, Robert Tibshirani, Ryan Tibshirani, Alexander Tsybakov, Sara Van de Geer, Larry Wasserman, Frank Werner, Bin Yu, Ming Yuan, and Harry Zhou. The Statistics and EECS departments, staff, faculty and students at UC Berkeley have provided a wonderful intellectual environment for this work; I also thank the Statistics Group at ETH Zurich, as well as the Laboratory for Information and Decision Sciences (LIDS) at MIT for their support during my visiting professor stays.

Over the years, I have also had the pleasure of working with many outstanding students and postdoctoral associates. Our discussions and research together as well as their feedback have been instrumental in shaping and improving this book. I would like to thank my current and former students, postdocs and visting fellows—many of whom are now current colleagues—including Alekh Agarwal, Arash Amini, Sivaraman Balakrishan, Merle Behr, Joseph Bradley, Yudong Chen, Alex Dimakis, John Duchi, Reinhard Heckel, Nhat Ho, Johannes Lederer, Po-Ling Loh, Sahand Negahban, Xuanlong Nguyen, Nima Noorshams, Jonas Peters, Mert Pilanci, Aaditya Ramdas, Garvesh Raskutti, Pradeep Ravikumar, Prasad Santhanam, Nihar Shah, Yuting Wei, Fanny Yang, Yun Yang, and Yuchen Zhang. Yuting Wei put her artistic skills to excellent use in designing the cover for this book. Last but not least, I would like to thank Cambridge University Press, and in particular the support and encouragement throughout the process of Senior Editor Diana Gillooly. My apologies to anyone whose help I may have failed inadvertently to acknowledge.