Mathematical Foundations of Infinite-Dimensional Statistical Models

In nonparametric and high-dimensional statistical models, the classical Gauss–Fisher–Le Cam theory of the optimality of maximum likelihood and Bayesian posterior inference does not apply, and new foundations and ideas have been developed in the past several decades. This book gives a coherent account of the statistical theory in infinite-dimensional parameter spaces. The mathematical foundations include self-contained 'mini-courses' on the theory of Gaussian and empirical processes, on approximation and wavelet theory and on the basic theory of function spaces. The theory of statistical inference in such models – hypothesis testing, estimation and confidence sets – is then presented within the minimax paradigm of decision theory. This includes the basic theory of convolution kernel and projection estimation, as well as Bayesian nonparametrics and nonparametric maximum likelihood estimation. In the final chapter, the theory of adaptive inference in nonparametric models is developed, including Lepski's method, wavelet thresholding and adaptive confidence regions for self-similar functions.

EVARIST GINÉ (1944–2015) was the head of the Department of Mathematics at the University of Connecticut. Giné was a distinguished mathematician who worked on mathematical statistics and probability in infinite dimensions. He was the author of two books and more than a hundred articles.

RICHARD NICKL is Reader in Mathematical Statistics in the Statistical Laboratory in the Department of Pure Mathematics and Mathematical Statistics at the University of Cambridge. He is also Fellow in Mathematics at Queens' College, Cambridge.

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 of the state of the art in classical methods. While emphasising 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:

14. *Statistical Analysis of Stochastic Processes in Time*, by J. K. Lindsey
15. *Measure Theory and Filtering*, by Lakhdar Aggoun and Robert Elliott
16. *Essentials of Statistical Inference*, by G. A. Young and R. L. Smith
17. *Elements of Distribution Theory*, by Thomas A. Severini
18. *Statistical Mechanics of Disordered Systems*, by Anton Bovier
19. *The Coordinate-Free Approach to Linear Models*, by Michael J. Wichura
20. *Random Graph Dynamics*, by Rick Durrett
21. *Networks*, by Peter Whittle
22. *Saddlepoint Approximations with Applications*, by Ronald W. Butler
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* (Fourth Edition), by Rick Durrett
32. *Stochastic Processes*, by Richard F. Bass
33. *Regression for Categorical Data*, by Gerhard Tutz
34. *Exercises in Probability* (Second Edition), by Loïc Chaumont and Marc Yor
35. *Statistical Principles for the Design of Experiments*, by R. Mead, S. G. Gilmour and A. Mead
36. *Quantum Stochastics*, by Mou-Hsiung Chang
37. *Nonparametric Estimation under Shape Constraints*, by Piet Groeneboom and Geurt Jongbloed
38. *Large Sample Covariance Matrices*, by Jianfeng Yao, Zhidong Bai and Shurong Zheng

Mathematical Foundations of Infinite-Dimensional Statistical Models

Evarist Giné

Richard Nickl
University of Cambridge

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

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/9781107043169

First published 2016

Printed in the United States of America

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

Library of Congress Cataloging in Publication Data
Giné, Evarist, 1944–2015
Mathematical foundations of infinite-dimensional statistical models /
Evarist Giné, University of Connecticut, Richard Nickl, University of Cambridge.
pages cm. – (Cambridge series in statistical and probabilistic mathematics)
Includes bibliographical references and index.
ISBN 978-1-107-04316-9 (hardback)
1. Nonparametric statistics. 2. Function spaces. I. Nickl, Richard, 1980– II. Title.
QA278.8.G56 2016
519.5′4–dc23 2015021997

ISBN 978-1-107-04316-9 Hardback

A la meva esposa Rosalind

Dem Andenken meiner Mutter Reingard, 1940–2010

Contents

Preface

The classical theory of statistics was developed for parametric models with *finite-dimensional parameter spaces*, building on fundamental ideas of C. F. Gauss, R. A. Fisher and L. Le Cam, among others. It has been successful in providing modern science with a paradigm for making statistical inferences, in particular, in the 'frequentist large sample size' scenario. A comprehensive account of the mathematical foundations of this classical theory is given in the monograph by A. van der Vaart, *Asymptotic Statistics* (Cambridge University Press, 1998).

The last three decades have seen the development of statistical models that are infinite (or 'high') dimensional. The principal target of statistical inference in these models is a function or an infinite vector f that itself is not modelled further parametrically. Hence, these models are often called, in some abuse of terminology, *nonparametric models*, although f itself clearly also is a parameter. In view of modern computational techniques, such models are tractable and in fact attractive in statistical practice. Moreover, a mathematical theory of such nonparametric models has emerged, originally driven by the Russian school in the early 1980s and since then followed by a phase of very high international activity.

This book is an attempt to describe some elements of the mathematical theory of statistical inference in such *nonparametric*, or infinite-dimensional, models. We will first establish the main probabilistic foundations: the theory of Gaussian and empirical processes, with an emphasis on the 'nonasymptotic concentration of measure' perspective on these areas, including the pathbreaking work by M. Talagrand and M. Ledoux on concentration inequalities for product measures. Moreover, since a thorough understanding of infinite-dimensional models requires a solid background in functional analysis and approximation theory, some of the most relevant results from these areas, particularly the theory of wavelets and of Besov spaces, will be developed from first principles in this book.

After these foundations have been laid, we turn to the statistical core of the book. Comparing nonparametric models in a very informal way with classical parametric models, one may think of them as models in which the number of parameters that one estimates from the observations is *growing proportionally to sample size n* and has to be carefully selected by the statistician, ideally in a data-driven way. In practice, nonparametric modelling is often driven by the honesty of admitting that the traditional assumption that n is large compared to the number of unknown parameters is too strong. From a mathematical point of view, the frequentist theory that validates statistical inferences in such models undergoes a radical shift: leaving the world of finite-dimensional statistical models behind implies that the likelihood function no longer provides 'automatically optimal' statistical methods ('maximum likelihood estimators') and that extreme care has to be exercised when constructing inference procedures. In particular, the Gauss–Fisher–Le Cam efficiency theory based on the Fisher information typically yields nothing informative about what optimal procedures are in nonparametric statistics, and a new theoretical framework is required. We will show how the minimax paradigm can serve as a benchmark by which a theory of optimality in nonparametric models can be developed. From this paradigm arises the 'adaptation' problem, whose solution has been perhaps one of the major achievements of the theory of nonparametric statistics and which will be presented here for nonparametric function estimation problems. Finally, likelihood-based procedures can be relevant in nonparametric models as well, particularly after some regularisation step that can be incorporated by adopting a 'Bayesian' approach or by imposing qualitative a priori shape constraints. How such approaches can be analysed mathematically also will be shown here.

Our presentation of the main statistical materials focusses on function estimation problems, such as density estimation or signal in white-noise models. Many other nonparametric models have similar features but are formally different. Our aim is to present a unified statistical theory for a canonical family of infinite-dimensional models, and this comes at the expense of the breadth of topics that could be covered. However, the mathematical mechanisms described here also can serve as guiding principles for many nonparametric problems not covered in this book.

Throughout this book, we assume familiarity with material from real and functional analysis, measure and probability theory on the level of a US graduate course on the subject. We refer to the monographs by G. Folland, *Real Analysis* (Wiley, 1999), and R. Dudley, *Real Analysis and Probability* (Cambridge University Press, 2002), for relevant background. Apart from this, the monograph is self-contained, with a few exceptions and 'starred sections' indicated in the text.

This book would not have been possible without the many colleagues and friends from whom we learnt, either in person or through their writings. Among them, we would like to thank P. Bickel, L. Birgé, S. Boucheron, L. Brown, T. Cai, I. Castillo, V. Chernozhukov, P. Dawid, L. Devroye, D. Donoho, R. Dudley, L. Dümbgen, U. Einmahl, X. Fernique, S. Ghosal, A. Goldenshluger, Y. Golubev, M. Hoffmann, I. Ibragimov, Y. Ingster, A. Iouditski, I. Johnstone, G. Kerkyacharian, R. Khasminskii, V. Koltchinskii, R. Latala, M. Ledoux, O. Lepski, M. Low, G. Lugosi, W. Madych, E. Mammen, D. Mason, P. Massart, M. Nussbaum, D. Picard, B. Pötscher, M. Reiß, P. Rigollet, Y. Ritov, R. Samworth, V. Spokoiny, M. Talagrand, A. Tsybakov, S. van de Geer, A. van der Vaart, H. van Zanten, J. Wellner, H. Zhou and J. Zinn.

We are grateful to A. Carpentier, I. Castillo, U. Einmahl, D. Gauthier, D. Heydecker, K. Ray, J. Söhl and B. Szabò for proofreading parts of the manuscript and providing helpful corrections.

Moreover, we are indebted to Diana Gillooly of Cambridge University Press for her support, patience and understanding in the process of this book project since 2011.

R.N. would also like to thank his friends N. Berestycki, C. Damböck, R. Dawid and M. Neuber for uniquely stimulating friendships that have played a large role in the intellectual development that led to this book (and beyond).

Outline and Reading Guide

In principle, all the chapters of this book can be read independently. In particular, the chapters on Gaussian and empirical processes, as well as the one on function spaces and approximation theory, are mostly self-contained. A reader interested primarily in the 'statistical chapters' (5 through 8) may choose to read those first and then turn to the mathematical foundations laid out in Chapters 2 through 4 later, when required. A short outline of the contents of each chapter is given in the following paragraphs:

Chapter 1 introduces the kinds of statistical models studied in this book. In particular, we will discuss why many common 'regular' regression models with normally distributed error terms can be mathematically accommodated within one Gaussian function estimation problem known as the *Gaussian white noise model*.

Chapters 2 and 3 lay the probabilistic foundations of much of the statistical theory that follows: one chapter on Gaussian processes and one on empirical processes. The Gaussian theory is mostly classical, presented with a focus on statistically relevant materials, such as the isoperimetric inequality for Gaussian measures and its consequences on concentration, as well as a study of suprema of Gaussian processes. The theory for empirical measures reflects the striking recent developments around the concentration-of-measure phenomenon. Effectively, here, the classical role of the central limit theorem in statistics is replaced by nonasymptotic concentration properties of product measures, as revealed in fundamental work by Talagrand, Ledoux, Massart and others. This is complemented by a treatment of abstract empirical process theory, including metric entropy methods, Vapnik-Červonenkis classes and uniform central limit theorems.

Chapter 4 develops from first principles some key aspects of approximation theory and its functional analytic foundations. In particular, we give an account of wavelet theory and of Besov spaces, with a focus on results that are relevant in subsequent chapters.

Chapter 5 introduces basic linear estimation techniques that are commonly used in nonparametric statistics, based on convolution kernels and finite-dimensional projection operators. Tools from Chapters 3 and 4 are used to derive a variety of probabilistic results about these estimators that will be useful in what follows.

Chapter 6 introduces a theoretical paradigm – the *minimax paradigm* – that can be used to objectively measure the performance of statistical methods in nonparametric models. The basic information-theoretic ideas behind it are developed, and it is shown how statistical inference procedures – estimators, tests and confidence sets – can be analysed and compared from a minimax point of view. For a variety of common nonparametric models, concrete constructions of minimax optimal procedures are given using the results from previous chapters.

Chapter 7 shows how the likelihood function can still serve as a successful guiding principle in certain nonparametric problems if a priori information is used carefully. This can be done by imposing certain qualitative constraints on the statistical model or by formally adopting a Bayesian approach which then can be analysed from a frequentist point of view. The key role of the Hellinger distance in this theory (as pointed out in work by Le Cam, Birgé, van de Geer, van der Vaart and others) is described in some detail.

Chapter 8 presents the solution to the nonparametric adaptation problem that arises from the minimax paradigm and gives a theory of statistical inference for ‘fully automatic’ statistical procedures that perform well over maximal collections of nonparametric statistical models. Surprising differences are shown to arise when considering the existence of adaptive estimation procedures in contrast to the existence of associated adaptive confidence sets. A resolution of this discrepancy can be obtained by considering certain nonparametric models of ‘self-similar’ functions, which are discussed in some detail and for which a unified theory of optimal statistical inference can be developed.

Each chapter is organised in several sections, and historical notes complementing each section can be found at the end of each chapter – these are by no means exhaustive and only indicate our understanding of the literature.

At the end of each section, exercises are provided: these, likewise, complement the main results of the text and often indicate interesting applications or extensions of the materials presented.

Postscript

It is a terrible tragedy that Evarist Giné passed away shortly after we completed the manuscript. His passion for mathematics was exceeded only by his love for his wife, Rosalind; his daughters, Núria and Roser; and his grandchildren, Liam and Mireia. He mentioned to me in September 2014, when I last met him in Cambridge (MA), that perhaps he wanted to dedicate this book to all of them, but in an e-mail to me in January 2015, he mentioned explicitly that he wanted it to be for Rosalind. I have honoured his decision; however, I know that with this last work he wanted to thank all of them for having been his wonderful family – who continue his infectious passion into new generations.

I am myself deeply grateful to my father, Harald, for all his support and inspiration throughout my life in all domains. I dedicate this book to the memory of my mother, Reingard, in loving gratitude for all her courage and everything she has done for me. And of course, insofar as this book relates to the future, it is for Ana and our son, Julian, with love and affection.