# Nonparametric Estimation under Shape Constraints

This book treats the latest developments in the theory of order-restricted inference, with special attention to nonparametric methods and algorithmic aspects. Among the topics treated are current status and interval censoring models, competing risk models and deconvolution. Methods of order-restricted inference are used in computing maximum likelihood estimators and developing distribution theory for inverse problems of this type.

The authors have been active in developing these tools and present the state of the art and open problems in the field. The earlier chapters provide an introduction to the subject, while the later chapters are written with graduate students and researchers in mathematical statistics in mind. Each chapter ends with a set of exercises of varying difficulty. The theory is illustrated with the analysis of real-life data, which are mostly medical in nature.

Piet Groeneboom is Professor Emeritus of Statistics at Delft University of Technology, the Netherlands.

Geurt Jongbloed is Professor of Statistics at Delft University of Technology, the Netherlands.

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:

13. *Exercises in Probability*, by Loïc Chaumont and Marc Yor
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. *Analysis of Multivariate and High-Dimensional Data*, by Inge Koch
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

Cambridge University Press & Assessment
978-0-521-86401-5 — Nonparametric Estimation under Shape Constraints
Piet Groeneboom , Geurt Jongbloed
Frontmatter
More Information

# Nonparametric Estimation under Shape Constraints

## Estimators, Algorithms and Asymptotics

Piet Groeneboom
*Delft University of Technology*

Geurt Jongbloed
*Delft University of Technology*

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, 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

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521864015



First published 2014

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

*Library of Congress Cataloging-in-Publication data*
Groeneboom, P.
Nonparametric estimation under shape constraints : estimators, algorithms and asymptotics / Piet Groeneboom,
Technische Universiteit Delft, The Netherlands, Geurt Jongbloed, Technische Universiteit Delft, The Netherlands.
pages cm. – (Cambridge series in statistical and probabilistic mathematics ; 38)
Includes bibliographical references and index.
ISBN 978-0-521-86401-5 (hbk.)
1. Nonparametric statistics. 2. Multivariate analysis. 3. Mathematical statistics. 4. Estimation theory.
I. Jongbloed, Geurt, 1968– II. Title.
QA278.8.G76 2014
519.5´4–dc23 2014022148

ISBN 978-0-521-86401-5 Hardback

Additional resources for this publication at http://statistics.tudelft.nl/CUPbook

# Contents

Cambridge University Press & Assessment
978-0-521-86401-5 — Nonparametric Estimation under Shape Constraints
Piet Groeneboom , Geurt Jongbloed
Frontmatter
More Information

# Preface and Acknowledgments

Research on nonparametric estimation under shape constraints started in the 1950s. Papers such as Ayer et al., 1955, and Van Eeden, 1956, appeared on estimation of functions under the restriction of monotonicity or unimodality, more generally called isotonic estimation. An isotonic estimator is an estimator that is computed under an order restriction, where the order can be a partial order. The order restriction can also be imposed on the derivative of the estimator, so an estimator of a convex function (in dimension one or higher), which is itself also convex, is also called an isotonic estimator.

A summary of the early work was given in the well-known book by Barlow et al., 1972, on isotonic regression. Originally, the focus was on defining and constructing estimators satisfying these order constraints. As an example, in Grenander, 1956, it is shown that the (nonparametric) maximum likelihood estimator (MLE) of a monotone decreasing density can be constructed as the left-continuous slope of the least concave majorant of the empirical distribution function. Developing asymptotic distribution theory for these isotonic estimators turned out to be rather difficult. Nonnormal limit distributions appear and rates of convergence are slower than the square root of the sample size. This behavior is now commonly classified as belonging to the area of nonstandard asymptotics. In the case of the mentioned Grenander MLE, the rate of convergence of this estimator (evaluated at a fixed point, under some local assumptions) is the cube root of the sample size. Moreover, the nonnormal asymptotic distribution of the estimator is (after rescaling) the so-called Chernoff distribution, which is (up to a factor 2) the distribution of the derivative of the greatest convex minorant of two-sided Brownian motion with parabolic drift, evaluated at zero.

Research on isotonic regression received new impetus in the 1990s when it became clear that it was the right setting for studying (nonparametric) MLEs of the distribution function in inverse problems. Examples of such problems include interval censoring models such as the current status model, deconvolution problems and the classical Wicksell corpuscle problem. Current status data or interval censored data are quite common in medical research, but are also relevant for econometric models such as the binary choice model. In the context of the current status model, the same (Chernoff) limit distribution appears for the MLE of the (by definition) monotone distribution function as for the Grenander estimator of a monotone density. Whether this Chernoff distribution also gives the (pointwise) limit behavior of the MLE of the distribution function in the more general interval censoring problem or a large class of deconvolution problems is still an open question. Very specific conjectures for the convergence to this limit distribution have been formulated, though.

As mentioned, the Chernoff distribution is the distribution of a functional of Brownian motion. Also, other functionals of Brownian motion appear in the limit theory for nonparametric estimators, for example, in the situation of estimating a convex function and its derivative at a fixed point. The local limit of a nonparametric least squares estimator of a smooth convex regression function can be characterized as the second derivative of an "invelope" of integrated Brownian motion plus the 4th power of the time variable, instead of Brownian motion with parabolic drift, which figured in the limit distribution of monotone estimators. The estimators have pointwise rates of convergence $n^{2/5}$ for the convex function and $n^{1/5}$ for its derivative, rather than the $n^{1/3}$ occurring in the situation of estimating a monotone function.

It is the purpose of this book to introduce the subject of shape-restricted statistical inference, to present the current state of the theory and also to describe still open problems. The subjects covered include those discussed in part 2 of the book by Groeneboom and Wellner, 1992. That book is still available in a Kindle edition, but a lot of theory has been developed since 1992. As an example, in applying the maximum likelihood theory in inverse (often medical) problems, the maximum likelihood estimator will usually be a piecewise constant jump function, so estimation of a hazard or density function is only possible after some kind of smoothing. Theory about this, and theory about the smoothed maximum likelihood estimator (SMLE) and the maximum smoothed likelihood estimator (MSLE), has only recently been developed and is discussed in the present book.

Another direction of considerable progress has been the analysis of so-called smooth functionals in inverse problems. Although the local rate of convergence of the MLE in the inverse problems is usually slower than $n^{1/2}$, there are often smooth functionals of moment type that can be estimated at rate $n^{1/2}$ with normal limit distributions. This theory, which is far from complete, is treated in this book. It depends on properties of solutions of certain integral equations, which in many situations do not have explicit solutions.

Also, theory has been developed for testing problems in models with shape constraints. Examples are the two- and $k$-sample tests based on interval censored data. In contrast with testing theory in the presence of right censored data, the theory of these problems has had a very slow start. The main challenge is to construct test statistics that have distributions that do not depend on the observation time distributions in the samples. Such tests are presented in this book and compared with tests not having this property. The techniques used here are different from both the theory used for the local limits and the smooth functional theory. Moreover, various forms of the bootstrap play a very important role in determining the critical values for these tests.

We also discuss confidence intervals for distribution functions, densities and hazards, constructed by bootstrap procedures. These are compared with intervals, based on plug-in estimators for the variance of the asymptotic distribution and intervals based on likelihood ratio tests, using the MLE (which only exist for distribution functions and not for densities or hazards). The confidence intervals also are computed for some real-life data sets, such as the Bangkok cohort data, which were kindly provided to us by the researchers in the Bangkok Metropolitan Administration Injecting Drug Users cohort study and Michael Hudgens (see the acknowledgments). We also constructed confidence intervals for the hepatitis A data, provided to us by Niels Keiding.

Throughout the years, we have written computer programs related to the subject of this book. We will make relevant programs (some of these need some rewriting) public via the website http://statistics.tudelft.nl/CUPbook.

The book can be used as the textbook for an advanced undergraduate course. Working knowledge of basic probability theory and mathematical statistics is assumed. Chapters 1 through 7 are rather general and focus on modeling of data and the derivation of nonparametric estimators for functions within these models. Parts of Chapter 3 could be skipped for an undergraduate course. Chapter 7 focuses on algorithms that can be used to compute the shape-constrained estimators in particular models. The later chapters are more technical and can be used as ingredients of a graduate course. These chapters can certainly be used to define concrete research projects in the area. Every chapter concludes with a number of exercises of varying levels. Some of these are meant to fill in details of arguments given in the text. Others extend results obtained in the text or present additional results.

Apart from being used as a textbook, we hope the book will also be used by colleagues and inspire them to work in the field of shape-restricted statistical inference. We also hope that people involved in medical statistics, econometrics and other fields of application will use the book to learn about shape-restricted statistics and benefit from the progress made in this field.

## Acknowledgments

We would like to thank Diana Gillooly for her support in the process of writing this book. Also we thank our co-authors of various papers in the area of shape-constrained estimation, including Lutz Dümbgen, Bert van Es, Ronald Geskus, Gerard Hooghiemstra, Vladimir Kulikov, Steven Lalley, Rik Lopuhaä, Marloes Maathuis, Frank van der Meulen, Stefanie Michael, Nico Temme, Jon Wellner and Birgit Witte.

We also wish to express our thanks to Niels Keiding for providing us with various data sets and Kachit Choopanya, Dwip Kitayaporn, Timothy Mastro, Philip Mock and Suphak Vanichseni for allowing us to use the data from the Bangkok Metropolitan Administration Injecting Drug Users cohort study.

Piet Groeneboom and Geurt Jongbloed
Delft, 2014