An Introduction to the Advanced Theory of Nonparametric Econometrics

Interest in nonparametric methodology has grown considerably over the past few decades, stemming in part from vast improvements in computer hardware and the availability of new software that allows practitioners to take full advantage of these numerically intensive methods. This book is written for advanced undergraduate students, intermediate graduate students, and faculty, and provides a complete teaching and learning course at a more accessible level of theoretical rigor than Racine's earlier book co-authored with Qi Li, Nonparametric Econometrics: Theory and Practice (2007). The open source R platform for statistical computing and graphics is used throughout in conjunction with the R package np. Recent developments in reproducible research is emphasized throughout with appendices devoted to helping the reader get up to speed with R, R Markdown, TeX and Git.

Jeffrey S. Racine is Professor in the Department of Economics and Professor in the Graduate Program in Statistics in the Department of Mathematics and Statistics at McMaster University, Canada. He holds the Senator William McMaster Chair in Econometrics and is a Fellow of the Journal of Econometrics. He is co-author of Nonparametric Econometrics: Theory and Practice (2007). He has published extensively in his field and has co-authored the R packages np and crs that are available on the Comprehensive R Archive Network (CRAN).

An Introduction to the Advanced Theory of Nonparametric Econometrics

A Replicable Approach Using R

JEFFREY S. RACINE
McMaster University, Ontario

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/9781108483407
DOI: 10.1017/9781108649841

First published 2019

Printed in the United States of America by Sheridan Books, Inc.

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

ISBN 978-1-108-48340-7 Hardback

Contents

Cambridge University Press
978-1-108-48340-7 — An Introduction to the Advanced Theory and Practice of Nonparametric Econometrics
Jeffrey S. Racine
Frontmatter
More Information

List of Tables

List of Figures

Preface

In the early 20th century, the pioneering statistician Sir R. A. Fisher (1890-1962) set in motion what is known today as the *classical parametric Fisherian* approach by casting statistical estimation as a problem involving a *finite* number of parameters. However, parametric models provide only an approximation to the underlying data generating process and may therefore be biased and inconsistent. Models that seek to describe the data generating process in a statistically consistent manner are more involved, since the unknown components in such models are functions that fully characterize the underlying joint distribution of a data sample. *Nonparametric* methods are suitable for the estimation of an unknown function that belongs to a very broadly defined class of functions, and in this context, the number of parameters involved is said to be of *infinite* dimension. Although the complexity of nonparametric estimators often exceeds that of their more rigid parametric counterparts, they offer practitioners alternative approaches that can reveal features present in a data sample that might otherwise remain undetected.

Interest in nonparametric methodology has grown considerably over the past few decades, stemming in part from vast improvements in computer hardware and the availability of new software that allows practitioners to take full advantage of these numerically intensive methods. The earliest work on nonparametric *kernel* estimation of *probability density functions* dates back to the early 1950s (Fix and Hodges, 1951), on kernel estimation of *regression functions* to the 1960s (Watson, 1964), and on kernel estimation of *probability mass functions* to the 1970s (Aitchison and Aitken, 1976). There exist a variety of books that are devoted to nonparametric estimation and inference, although most of them appear to have been written with an audience of advanced graduate students and researchers in mind, and their focus is often on one very specific aspect of the field (e.g., density estimation). A list of notable contributions would include

- Prakasa Rao (1983; Prakasa Rao, 2014) (devoted to large sample properties of various nonparametric estimators)
- Devroye and Györfi (1985) (devoted to the L_1 approach to nonparametric estimation)

- Silverman (1986) (devoted to density estimation and related topics)
- Härdle (1990) (devoted to applied nonparametric regression)
- Scott (1992) (devoted to density estimation and high-dimensional visualization)
- Wand and Jones (1995) (devoted to an accessible treatment of kernel density estimation and regression)
- Fan and Gijbels (1996) (devoted to local polynomial estimation)
- Simonoff (1996) (devoted to smooth density estimation, regression, and ordered categorical data)
- Bowman and Azzalini (1997) (devoted to the application of kernel methods in S-plus)
- Hart (1997) (devoted to nonparametric smoothing and lack-of-fit tests)
- Bosq (1998) (devoted to the theory of kernel methods for dependent data)
- Horowitz (1998) (devoted to semiparametric econometric methods)
- Pagan and Ullah (1999) (first broad treatment of nonparametric econometrics)
- Fan and Yao (2003) (devoted to time series modeling)
- Yatchew (2003) (devoted to applied semiparametric methods using a differencing technique)
- Ruppert et al. (2003) (devoted to semiparametric modeling)
- Härdle et al. (2004) (devoted to nonparametric and semiparametric modeling)
- Wasserman (2006) (devoted to brief accounts of many modern topics in nonparametric inference)
- Li and Racine (2007) (devoted to nonparametric and semiparametric modeling with an emphasis on categorical covariates)
- Tsybakov (2009) (devoted to construction of optimal estimators, minimax optimality and adaptivity)
- Ahamada and Flachaire (2010) (devoted to an accessible introduction to nonparametric and semiparametric econometrics)
- Henderson and Parmeter (2015) (devoted to an accessible treatment of nonparametric econometrics)
- Politis (2015) (devoted to a transformation-based approach to model free inference)
- Hansen (2018) (devoted to econometrics but with chapters for kernel regression and density estimation)

In Li and Racine (2007), our aim was to provide a rigorous and comprehensive treatment of nonparametric econometric methodology, with an emphasis on mixed categorical and continuous data settings, intended for advanced graduate students and researchers looking to keep abreast of this rapidly growing field. The accompanying R (R Core Team, 2018) package, titled np, (Hayfield and Racine, 2008) was intended to facilitate the implementation in applied research settings of many of the methods that we discussed. We are grateful for the constructive criticism and helpful feedback that we have received about these projects, and we owe an enormous debt to the scores of researchers whose work made them possible.

In this book, we are aiming our attention squarely at advanced undergraduate students, intermediate graduate students, and faculty who wish to explore this exciting field, although not necessarily at the level of theoretical rigour that was found in our previous treatment. We take a more *organic* approach than existing treatments of the subject, and present a unique sequence of topics that are not collectively found elsewhere. We begin with a simple estimator that is standard fare in introductory statistics courses, namely the sample proportion, which is a nonsmooth nonparametric estimator of an unknown probability. This serves as preliminary motivation for the progressive introduction of kernel-smoothing, density estimation, conditional density estimation, and the estimation of more general conditional moments such as the conditional mean (regression), variance, and related objects. Proof concepts are illustrated *once* when each unique case is first encountered, whereas proofs that are of a similar nature to those already treated are either relegated to exercises or accompanied by citation info so that the interested reader may find them in existing treatments. Our approach emphasizes the plug-in principle that is the essence of most nonparametric methods. This involves identifying a fundamental statistical object (e.g., a conditional mean), expressing the object in terms of unknown density or distribution functions, and then plugging in *smooth* and *consistent* estimates of these unknowns. Special attention is also given to smoothing parameter selection and to the statistical properties of the estimator that results.

Our treatment of nonparametric estimation evolves along the lines of what one might encounter in an introductory statistics course, closely following the conventional sequence of topics. That convention is to first introduce discrete probability (i.e., mass) functions in Chapter 1 and then proceed to the study of continuous probability density functions in Chapter 2. However, one chapter that is conspicuously absent from introductory courses is a chapter on probability distributions with mixed discrete and continuous features (such problems are known to be “parametrically awkward” (Aitchison and Aitken, 1976, page 419)). In a nonparametric framework, modeling such objects isn’t awkward at all, and hence we fill this gap in Chapter 3 with a treatment of mixed discrete and continuous probability density functions and their cumulative counterparts. Moreover, it will be seen that we can subsequently tackle in a seamless manner *any* statistical object that is defined over mixed discrete and continuous data. Along the way, we will also cover nonparametric estimation of smooth quantile functions and copula functions. We then consolidate and fix notation by means of a parsimonious representation of the mixed-data multivariate product kernel. This then allows us to plunge into a range of methods for estimation and inference including nonparametric regression, nonparametric modeling of volatility, as well as methods for stationary time series.

We assess *pointwise* and *global* estimation error via the mean square and integrated (summed) mean square error, respectively. The pointwise error of estimation at a given point x is the difference between an estimate of the statistical object of interest and the object itself. For instance, we might compute the difference between the empirical CDF $F_n(x)$ and the unknown CDF $F(x)$. Pointwise error is a simple measure that is useful for the construction of confidence intervals. The *integrated* mean square error (or the *summed* mean square error in the context of discrete support random variables) measures the overall error of estimation and is useful as a criterion for bandwidth selection. *Uniform* error is another metric that is computed as the maximal difference between the estimate and the object, i.e., $\sup_x |F_n(x) - F(x)|$. It is typically approached using empirical process theory (Prakasa Rao, 2014). Uniform error is useful for placing bounds on other types of error and establishing *simultaneous* or uniform confidence bands. In this book, we consider only the first two types of error (pointwise and global) and direct the reader whose interest lies in uniform error to other more advanced treatments.

We emphasize how kernel estimators can be interpreted as *shrinkage* estimators (Stein, 1956), as demonstrated in Kiefer and Racine (2009) and Kiefer and Racine (2017). From this perspective, the local constant, local linear, and other variants of local polynomial kernel estimators can be improved; for a broad class of data-generating processes (the class of *analytic* functions), these estimators are able to achieve the rate of convergence that is associated with correctly specified parametric models. Theoretical underpinnings for this result, which is achieved through joint selection of the polynomial degree and bandwidth vectors, can be found in Hall and Racine (2015). Although this approach requires a solution to a mixed-integer problem, its implementation is now feasible in R, and this represents an exciting advance in the area of local polynomial estimation of statistical objects. The interpretation of kernel methods as shrinkage estimators is underscored wherever appropriate in each chapter. Simulations and practical exercises reveal that the performance of this estimator may be superior to that of alternative approaches that are based on ad hoc selection of the polynomial order. Our perspective on kernel estimators, as seen through the lens of shrinkage estimators, is quite novel and, to the best of our knowledge, is not found elsewhere.

The computational run time of various routines in the R package np (Hayfield and Racine, 2008) can be reduced through their ability to exploit the power of multiple processors (see the R package npRmpi) and through their incorporation of algorithmic enhancements such as the use of trees. That being said, kernel methods are computationally intensive relative to many of their parametric peers; however, patience in this regard often pays dividends.

R code for all examples in this book is sourced from an R Markdown script and can be studied and modified by readers (this document is composed in R Markdown and uses R bookdown extensions (https://bookdown.org/yihui/bookdown) (Xie, 2017)). Each chapter ends with a *Practitioner's Corner* that provides a set of commented examples in R that can be refined by the reader to suit their needs. A solutions manual is available to instructors along with LaTeX *Beamer* PDF formatted slides authored in R Markdown that can be modified and tailored to an instructor's needs.

In this book, we derive results only for the *notationally parsimonious* case involving univariate data (or univariate conditioning/conditioned variables). Where appropriate, we present results for the multivariate case and draw attention to the salient differences between the two; however, for a thorough theoretical treatment of the multivariate cases, we simply direct the interested readers to Li and Racine (2007) and other sources. It is our conjecture that essentially all of the intuition underlying nonparametric kernel methods can be distilled from the univariate case, at least from the theoretical perspective. However, from the applied perspective, we impose no such limitations, and emphasize cases involving multivariate (and often mixed multivariate) data throughout.

We also touch upon a number of practical aspects of nonparametric kernel methods such as *kernel carpentry* (i.e., the construction of kernel functions with certain useful properties), and provide empirical examples to illustrate these concepts. We encourage the use of tools that facilitate reproducible research.

This book would not exist without the legacy (and ongoing) contributions of an incredibly talented global network of academics harbouring a wide array of research interests in the field of nonparametric statistics and econometrics. If you are reading this and have contributed to this exciting field, please take a virtual bow and accept our heartfelt thanks.

I would like to thank an abbreviated cast of characters, without whom this project would not exist. Qi Li, a co-author on a range of projects, has been an ongoing source of guidance, support, and encouragement. Tristen Hayfield and Zhenghua Nie, co-authors on the R packages np and crs, respectively, have helped craft user-friendly and computationally efficient implementations of the procedures that are detailed in this book. Nick Kiefer, a co-author, was the first to open my eyes to the interpretation of kernel estimators as shrinkage estimators. Peter Hall, a co-author whose acumen, friendship, and wisdom are sorely missed, made enduring contributions to the field and left a rich legacy that will surely last for generations. I would also like to thank but not implicate John Kealey, a former Ph.D. student who painstakingly pored through this book and polished its many rough edges, along with the students and faculty who attended a graduate course at McMaster University in the Fall of 2017 and who reported numerous typos in early drafts of this book (Alyssa, Anthony, Camille, Francis, James, Joaquin, Karen, Mark, Richard, Yuyan, and Zvezdomir). And last but certainly not least, I am indebted to my wife Jennifer and son Adam, who endured far too many months of my seven-day-a-week obsession with this project.

This book is dedicated to the memory of our kind, gentle, generous, and irreplaceable colleague, Peter Gavin Hall AO FAA FRS (November 21, 1951—January 9, 2016), an Australian researcher who worked in the areas of probability theory and mathematical statistics. Peter was described by the American Statistical Association as one of the most influential and prolific theoretical statisticians in the history of the field. It is fitting that The School of Mathematics and Statistics Building at The University of Melbourne was renamed the Peter Hall Building in his honour on December 9, 2016.

Glossary of Notation

Object	Brief Definition
$\beta(x)$	marginal effects function (derivative or finite difference of $g(x)$)
$\hat{\beta}(x)$	kernel smoothed marginal effects function (derivative or finite difference of $\hat{g}(x)$)
$C(u_x, u_y)$	bivariate copula function
$f(x)$	probability density function
$\hat{f}(x)$	kernel smoothed probability density function
$f(y\|x)$	conditional probability density function
$\hat{f}(y\|x)$	kernel smoothed conditional probability density function
$F(x)$	cumulative distribution function
$F_n(x)$	empirical cumulative distribution function
$\hat{F}(x)$	kernel smoothed cumulative distribution function
$F(y\|x)$	cumulative conditional distribution function
$\hat{F}(y\|x)$	kernel smoothed cumulative conditional distribution function
γ	vector of bandwidths and smoothing parameters for q continuous, r unordered, and s ordered covariates
$G((x - X_i)/h)$	continuous support univariate cumulative probability density kernel function
$G_\gamma(X_i, x)$	mixed-data multivariate cumulative probability density kernel function
$g(x)$	conditional mean function
$\hat{g}(x)$	kernel smoothed conditional mean function
h	bandwidth for continuous covariate
$K((x - X_i)/h)$	continuous support univariate probability density kernel function
$K_\gamma(X_i, x)$	mixed-data multivariate probability density kernel function
λ	smoothing parameter for discrete covariate

Object	Brief Definition
$l(X_i, x, \lambda)$	unordered discrete support univariate probability mass kernel function
$L(X_i, x, \lambda)$	ordered discrete support univariate probability mass kernel function
$\mathcal{L}(X_i, x, \lambda)$	ordered discrete support univariate cumulative probability mass kernel function
$M(x)$	conditional mode function
$\hat{M}(x)$	kernel smoothed conditional mode function
$p(x)$	probability mass function
$p_n(x)$	empirical probability mass function (sample proportion)
$\hat{p}(x)$	kernel smoothed probability mass function
q_τ	unconditional quantile function (inverse CDF)
$\hat{q}_\tau$	kernel smoothed unconditional quantile function
$q_\tau(x)$	conditional quantile function (inverse conditional CDF)
$\hat{q}_\tau(x)$	kernel smoothed conditional quantile function